THE COST OF USEFUL NATURAL GRADIENT UPDATES

Anonymous authors
Paper under double-blind review

ABSTRACT

What information is needed to turn a natural-gradient direction into a useful finite update? Under a population Kullback–Leibler (KL) budget, we call a step *useful* if it is feasible and loses at most a fraction ε of the best feasible gain along the direction. We construct a four-state exponential family whose laws share their initial gradient, scalar Fisher information and natural gradient, yet two laws have disjoint useful-step sets. With these quantities supplied exactly and the law otherwise known only through draws, the family's worst-case sample complexity is $\Theta(\log(1/\delta)/(p\varepsilon^2))$ for small ε, where p scales rare-state probabilities and δ is the failure probability. The budget is fixed and the optimal gain stays bounded away from zero, so the step length, not the direction, carries this cost. For succinctly described event-tilt models, returning a useful step is NP-hard even with the exact natural gradient and efficient exact sampling. Recovering the unit natural gradient to constant error is also NP-hard even in a two-parameter logistic family with Fisher condition number at most 3. We also give matching sample bounds for event tilts, sample bounds for damped Fisher solves and a population-KL certificate for affine classifiers. In frozen-feature classifier heads, stopping at a sampled KL boundary succeeds in about half of the trials, and a 10% KL margin raises joint success above 93% at a KL budget of 0.01. Thus, knowing where to move is not enough: how far to move can carry an update's entire cost.

1 INTRODUCTION

An exact natural-gradient direction need not determine a useful finite update, which also needs a step length. Natural-gradient and trust-region methods estimate the update direction with care and usually take its length from the local quadratic model (Kakade, 2001; Schulman et al., 2015; Martens & Grosse, 2015).

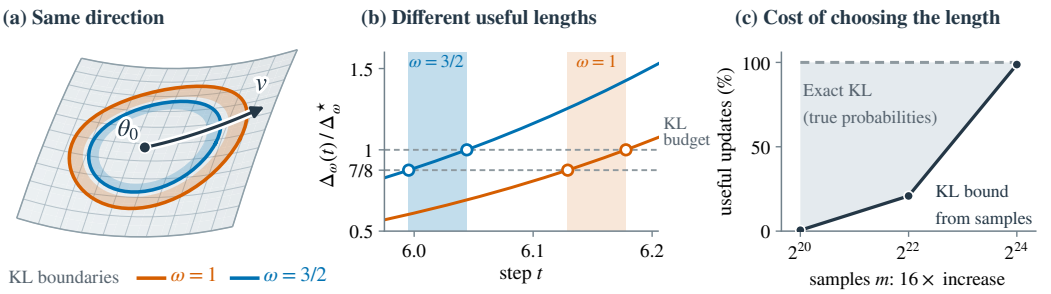

Figure 1: **Direction, calibration and sampling cost. (a)** Schematic: in the four-state laws P_ω of Theorem 3.2, ω scales the rare top state's probability. At $\omega = 1$ and $\omega = 3/2$, they share the gradient, Fisher matrix and natural gradient v at the initial parameters θ_0, but their KL boundaries (curves) and useful steps (shaded rings) differ. **(b)** Exact gain $\Delta_\omega(t)$ of the step t along v over the best feasible gain $\Delta^\star_\omega$, reached at the KL budget. Steps are useful where this ratio lies in $[7/8, 1]$ (shaded), and the two intervals are disjoint. **(c)** Logistic controls with directions estimated from m samples. A useful step respects the budget and attains at least 90% of the globally optimal KL-feasible query gain. Exact KL calibration is useful in every trial. A 95% confidence rule reusing the samples is useful in 0.5% to 98.7% of trials.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Under a population Kullback–Leibler (KL) budget, the step length can be far off even along the exact direction. In our four-state family, the quadratic step is over 450 times longer than any feasible step, and its KL is about 8.6×10^3 nats against a budget of $1/10$ (Section 3.1). Fisher approximation and inversion address direction recovery (Martens & Grosse, 2015; Martens, 2020), and step-selection analyses assume step-size or accuracy conditions (Section 2). We instead ask what information a useful length requires beyond an exact direction, and what acquiring it costs, a question that remains under-explored. We prove that knowing an *exact direction*, even with *efficient exact sampling*, does not make a useful finite step cheap to find. Here the cost comes from rare states rather than from dimension or conditioning, so better Fisher approximations cannot remove it, and the sample lower bound holds for *every procedure*, including line searches and backtracking.

At initial parameters $\theta_0 \in \mathbb{R}^d$, the natural gradient $v = F^{-1}b$ combines the nonzero objective gradient b and positive definite Fisher matrix F (Amari, 1998; Martens, 2020). For a KL budget $\eta > 0$, the quadratic trust-region model maximizes $b^\mathsf{T}\xi$ over parameter displacements ξ subject to $\frac{1}{2}\xi^\mathsf{T}F\xi \le \eta$. Its optimizer is $\xi = \sqrt{\frac{2\eta}{b^\mathsf{T}F^{-1}b}}\, v$. Let $\mathcal{K}(u)$ and $\Delta(u)$ denote the actual population KL cost and gain of parameters u. Over step lengths $t \ge 0$ along v, the best feasible gain is

$$\Delta^\star := \sup_{t \ge 0,\ \mathcal{K}(\theta_0 + tv) \le \eta} \Delta(\theta_0 + tv). \tag{1}$$

For $0 < \varepsilon < 1$, a step is *useful* if it stays within budget and attains at least $(1-\varepsilon)\Delta^\star$. *Calibration* selects such a length. A zero step fails the gain requirement whenever $\Delta^\star > 0$.

Local information does not determine the length. Our four-state family is a softmax distribution over statistic values $0, 1, 2, 3$, with their mean as objective, and a natural-gradient step shifts the logits along these values. The outcomes can be the actions of a policy at one state, where the best action is rarely chosen and the probabilities are known only through draws. At a fixed scale p_M of the rare-state probabilities, all laws in the family share the initial gradient, scalar Fisher information and natural gradient, yet some pairs of laws have disjoint intervals of useful steps (Figure 1(b)). Four fixed statistic values are the fewest that allow distinct laws with equal first two moments (Section 3.1).

Calibration costs as much as estimating a rare probability. With exact local information, a two-point lower bound and a matching confidence-calibration upper bound show that the sample complexity $m^\star(p_\mathrm{M}, \varepsilon, \delta)$, the fewest draws that yield a useful step with probability at least $1-\delta$, is $\Theta(\log(1/\delta)/(p_\mathrm{M}\varepsilon^2))$, as for estimating a rare probability (Dagum et al., 2000). The rare best outcome's tilted weight grows exponentially, so it can control finite-step KL while contributing little to the initial Fisher information. With the budget held fixed, the optimal gain stays above a positive constant as p_M decreases, so this cost does not come from a vanishing target.

Contributions. (i) Matching upper and lower bounds characterize calibration in the four-state family with exact local information (Theorem 3.2). (ii) For event tilts, an explicit confidence rule attains matching sample bounds, and selecting a useful tilt parameter from a succinct description is NP-hard, although the exact natural gradient is supplied and exact sampling is efficient (Theorem 3.3 and Corollary 3.4). (iii) Recovering the unit natural gradient of a two-parameter logistic model is NP-hard from its description, although its Fisher matrix has condition number at most 3, and needs exponentially many draws when only samples are available (Theorem 3.5). (iv) For bounded statistics, our sample bounds for estimating the damped natural gradient match up to a logarithmic factor in the dimension, at small damping and fixed confidence (Proposition 3.6). (v) For affine classifiers with known feature bounds, independent calibration draws certify the KL feasibility of a fixed set of candidate steps, even when the chosen step depends on those draws (Section 3.4 and Proposition F.1). (vi) Experiments match these predictions (Section 4). With m draws, count-simulation success at fixed $mp\varepsilon^2$ is similar across event probabilities p, exact-KL calibration meets the logistic gain target despite inaccurate directions, and a 10% KL margin lifts joint success in classifier heads from about 50% to over 93%.

2 Background & Related Work

Preliminaries. A *law* is a probability distribution. On a finite state space $\mathcal{S}$, the *base law* P describes the model before updating, and a *draw* is an independent exact sample from it unless stated otherwise. For statistics $h : \mathcal{S} \to \mathbb{R}^d$, exponential tilting defines $P_\theta(s) = e^{\langle \theta, h(s) \rangle} P(s)/Z(\theta)$, where

$Z(\theta) = \mathbb{E}_P e^{\langle \theta, h \rangle}$ (Wainwright & Jordan, 2008). The *reference law* for the KL budget is P, so an update costs $\mathrm{KL}(P\|P_\theta) = \mathbb{E}_P \log(P/P_\theta)$ nats. Substituting the tilted density gives $\mathrm{KL}(P\|P_\theta) = \log Z(\theta) - \langle \theta, \mathbb{E}_P h \rangle$. The Fisher information matrix $F(\theta) = (g_{ij}(\theta))$, with $g_{ij}(\theta) = \mathrm{Cov}_{P_\theta}(h_i, h_j)$, determines the quadratic term of KL between nearby parameters (Amari, 1982). For $d = 1$, it is the scalar Fisher information. We write $\mathbf{1}\{E\}$ for the indicator of a condition E and $\mathbf{1}_A(s) = \mathbf{1}\{s \in A\}$. For conditional models, population KL averages predictive KL over a reference input law. The population Fisher matrix averages outer products of the score over inputs and model-generated labels, whereas the empirical Fisher matrix uses observed labels (Kunstner et al., 2019).

A primer on complexity theory. Our results bound two resources, *draws* and *computation*, with one tool for each. A *procedure* is a randomized algorithm that uses an independent seed and at most a fixed number of draws, and returns a measurable output defined for every seed and draw sequence (Appendix A). Every step-selection rule in use is a procedure, from quadratic-model rescaling to TRPO's backtracking line search. The *sample complexity* of a problem is the smallest number of draws with which some procedure attains the required success probability under every law considered. *Sample lower bounds* allow unlimited computation and use *two-point testing* (Tsybakov, 2009). If two laws admit no common valid output, any procedure that succeeds under both must tell them apart from its draws, which takes more draws the closer the laws are in per-draw KL (Lemma A.1). *Computational lower bounds* start from a *succinct description*, here a Boolean formula φ with encoding $\langle \varphi \rangle$ that specifies the law. *Boolean satisfiability* (SAT) asks whether φ is satisfiable, and $\#\mathrm{SAT}(\varphi)$ counts its satisfying assignments (Valiant, 1979). A problem is *NP-hard* if a polynomial-time algorithm for it would decide SAT, and hence every problem in *NP* (nondeterministic polynomial time). A *language* is a set of binary strings, and a language in NP is *NP-complete* if every language in NP reduces to it by a polynomial-time *many-one reduction*, which maps each input to one with the same answer. *BPP* (bounded-error probabilistic polynomial time) contains problems that randomized polynomial-time algorithms solve with error at most $1/3$ (Arora & Barak, 2009). Each reduction decides SAT from one valid output, so a randomized polynomial-time procedure succeeding with probability at least $2/3$ on every input would imply $\mathrm{NP} \subseteq \mathrm{BPP}$, which is believed false.

Inference and sampling access. Evaluating the expected Fisher matrix involves inference in Bayesian networks (Kontkanen et al., 2000), where exact and relative-error inference are hard (Cooper, 1990; Dagum & Luby, 1993). Compiling a network into an arithmetic circuit makes inference linear in the circuit's size, but this size can be exponential in the size of the network's description (Darwiche, 2003). Our reductions allow efficient exact sampling and target the returned direction or step itself.

Step selection. Natural-policy-gradient analyses establish convergence with inexact policy evaluation (Agarwal et al., 2021; Cen et al., 2022). Conservative policy iteration (Kakade & Langford, 2002) and adaptive policy-gradient step sizes (Pirotta et al., 2013) maximize lower bounds on improvement. Probabilistic (Mahsereci & Hennig, 2015) and stochastic (Vaswani et al., 2019) line searches use noisy function values and gradients. Stochastic trust-region (Chen et al., 2018; Blanchet et al., 2019) and line-search analyses (Paquette & Scheinberg, 2020) impose probabilistic accuracy conditions. Jin et al. (2025) quantify oracle and sample costs for such methods.

Rare events and certification. Estimating a rare Bernoulli mass p to relative error ε at failure probability δ has classical sample order $\log(1/\delta)/(p\varepsilon^2)$ (Dagum et al., 2000). Our construction transfers this estimation cost to calibration. Our matching upper bounds control relative error using multiplicative concentration (Hoeffding, 1963) and Bernoulli-KL confidence intervals (Garivier & Cappé, 2011). Bu et al. (2018) study the related problem of estimating KL between unknown discrete laws. The equivalence of approximate counting and almost-uniform generation for self-reducible problems (Jerrum et al., 1986) concerns sampling solutions. Our reductions instead sample uniformly from all assignments, among which satisfying ones can be exponentially rare. Empirical Bernstein bounds (Maurer & Pontil, 2009) certify population KL using its observed per-input variance and known range. High-confidence policy improvement (Thomas et al., 2015) likewise uses finite-sample confidence bounds, but to certify a new policy's value rather than an update's KL.

3 Complexity of Useful Updates

After fixing what makes an update useful, we ask whether exact local information (Section 3.1), efficient exact sampling (Section 3.2) or a well-conditioned Fisher matrix (Section 3.3) makes a useful step or direction cheap to find. The answers are negative in the worst case, through sample

lower bounds that hold for every procedure, NP-hardness or both. Section 3.4 then shows what draws can still guarantee.

Definition 3.1 (Useful update and calibration)**.** Fix initial parameters $\theta_0 \in \mathbb{R}^d$, an objective J, a reference population and a declared update family $\mathcal{U} \subseteq \mathbb{R}^d$. An example is the ray $\{\theta_0 + tv : t \geq 0\}$ for a fixed direction $v \in \mathbb{R}^d$. For updated parameters $u \in \mathcal{U}$, the *KL cost* $\mathcal{K}(u)$ is the population KL divergence from the initial model to the updated one, and the *gain* is $\Delta(u) = J(u) - J(\theta_0)$. Given a budget $\eta > 0$, let $\Delta^\star = \sup\{\Delta(u) : u \in \mathcal{U},\ \mathcal{K}(u) \leq \eta\}$. For $0 < \varepsilon < 1$, an update u is *useful* if $\mathcal{K}(u) \leq \eta$ and $\Delta(u) \geq (1-\varepsilon)\Delta^\star$. *Calibration* is the problem of selecting a useful update from the available information. On the ray above, it seeks a feasible $(1-\varepsilon)$-approximation to Equation (1). *Local information* consists of the objective gradient, Fisher information matrix and natural gradient at θ_0. An *access model* specifies which of these are supplied exactly, and whether draws or a succinct description of the base law are available.

3.1 Calibration complexity with exact gradient and scalar Fisher information

We first ask whether exact local information determines a useful length. To test this, we look for laws that agree on every quantity in the local linear objective and quadratic KL models (Martens, 2020). When the objective is the mean of a statistic h, its gradient and scalar Fisher information both equal the variance of h, so such laws share the first two moments of h. The family below meets this constraint with four states and one free parameter ω, which moves mass to a rare top state. The paragraph after the theorem explains each constant.

Theorem 3.2 (Calibration with exact gradient and scalar Fisher information)**.** *Fix $0 < p_{\mathrm{M}} \leq 2^{-30}$, $0 < \varepsilon \leq 1/8$ and $0 < \delta \leq 1/3$. For an unknown $\omega \in [1, 3/2]$, let $h \in \{0,1,2,3\}$ have law*

$$P_\omega = \big(1 - (5+\omega)p_{\mathrm{M}},\ (3\omega - 2)p_{\mathrm{M}},\ (7 - 3\omega)p_{\mathrm{M}},\ \omega p_{\mathrm{M}}\big). \tag{2}$$

For $t \geq 0$, define $Z_\omega(t) = \mathbb{E}_{P_\omega} e^{th}$, $P_{\omega,t}(h) = e^{th}P_\omega(h)/Z_\omega(t)$, $J_\omega(t) = \mathbb{E}_{P_{\omega,t}} h$ and scalar Fisher information $F_\omega(t) = \mathrm{Var}_{P_{\omega,t}} h$. The gain and KL cost are $\Delta_\omega(t) = J_\omega(t) - 12p_{\mathrm{M}}$ and $\mathcal{K}_\omega(t) = \mathrm{KL}(P_\omega \| P_{\omega,t}) = \log Z_\omega(t) - 12p_{\mathrm{M}}t$, and the best feasible gain is $\Delta^\star_\omega = \max_{t \geq 0:\, \mathcal{K}_\omega(t) \leq 1/10} \Delta_\omega(t)$. Every law has $\mathbb{E}h = 12p_{\mathrm{M}}$ and $\mathbb{E}h^2 = 26p_{\mathrm{M}}$. Its initial gradient and Fisher information satisfy $J'_\omega(0) = F_\omega(0) = 26p_{\mathrm{M}} - 144p_{\mathrm{M}}^2 > 0$, and the natural gradient $F_\omega(t)^{-1}J'_\omega(t)$ is 1 for all $t \geq 0$. A procedure receives p_{M}, this family, h, the exact initial gradient and scalar Fisher information, and the exact natural gradient. Its only further information about ω is draws from P_ω. It must return a scalar $\widehat{t} \geq 0$ satisfying $\mathcal{K}_\omega(\widehat{t}) \leq 1/10$ and $\Delta_\omega(\widehat{t}) \geq (1-\varepsilon)\Delta^\star_\omega$. Requiring success with probability at least $1 - \delta$ for every $\omega \in [1, 3/2]$ gives sample complexity $\Theta(\log(1/\delta)/(p_{\mathrm{M}}\varepsilon^2))$.

Proof sketch. The candidate $t = \frac{1}{3}\log(1/(10\omega p_{\mathrm{M}}))$ gives the rare top state unnormalized weight $1/10$. Bounding the other tilted weights proves feasibility and $\Delta^\star_\omega > 27/100$, uniformly over the family. Gain and KL increase for $t > 0$. Useful steps form $[L_{\omega,\varepsilon}, T_\omega]$, where T_ω exhausts the KL budget and $L_{\omega,\varepsilon}$ attains the required gain fraction. The identity $Z_{\omega'}(t) - Z_\omega(t) = (\omega' - \omega)p_{\mathrm{M}}(e^t - 1)^3$ shows that the normalizers agree through second order at zero, but KL increases with ω for $t > 0$. Hence T_ω decreases. On each useful interval, the bounds $J'_\omega(t) \geq 3/5$ and $\Delta^\star_\omega \leq 7/20$ give $T_\omega - L_{\omega,\varepsilon} \leq 7\varepsilon/12$ by integration. Changing ω from 1 to $1 + 4\varepsilon$ moves the KL boundary by at least $4\varepsilon/5$, exceeding this width. Hence $T_{1+4\varepsilon} < L_{1,\varepsilon}$, so the intervals are disjoint. Thresholding a useful output therefore distinguishes two laws with identical local information and one-draw KL $O(p_{\mathrm{M}}\varepsilon^2)$, requiring $\Omega(\log(1/\delta)/(p_{\mathrm{M}}\varepsilon^2))$ draws. For the upper bound, the frequency of $h = 3$ gives an upper confidence estimate of ω. On the coverage event, its KL boundary is no larger than the true one. At the matching sample size, concentration controls coverage and estimation error, while boundary sensitivity bounds the lost gain by $\varepsilon\Delta^\star_\omega$. Appendix B.1 gives the proofs. □

Why four states and these constants? On at most three fixed, distinct statistic values, normalization and the first two moments determine the law, since the corresponding Vandermonde system has full column rank. On $\{0,1,2,3\}$, they leave one free direction, $p_{\mathrm{M}}(-1, 3, -3, 1)$, the third finite difference, which is orthogonal to 1, h and h^2. Varying ω moves P_ω along it, so every law has the same first two moments and local quantities, while the rare top state's mass ωp_{M} grows with ω. Its tilted weight dominates finite-step KL, so ω shifts the KL boundary and the useful steps. The range $\omega \in [1, 3/2]$ keeps every rare mass between p_{M} and $4p_{\mathrm{M}}$, and $\varepsilon \leq 1/8$ keeps the hard pair $\omega = 1$ and $\omega = 1 + 4\varepsilon$ inside it. The scale $p_{\mathrm{M}} \leq 2^{-30}$ keeps the middle states' tilted weights negligible

near the KL boundary, and the fixed budget $1/10$ keeps the target from shrinking with p_{M}. Both values only make the constants explicit, and $\delta \le 1/3$ is the usual regime for testing bounds. Beyond the construction, the quadratic KL model (Amari, 1998; Martens, 2020) behind trust-region steps (Schulman et al., 2015) uses only the statistic's covariance, whereas finite-step KL depends on its whole distribution. Any law with positive mass on at least four values of a scalar statistic has nearby laws with the same first two moments but different finite-step KL.

Consequence for local step rules. A rule using only the supplied local information cannot depend on ω. Each returned step is useful under at most one law of each hard pair. At $p_{\mathrm{M}} = 2^{-30}$, the quadratic-model length $\sqrt{\dfrac{2\eta}{b^{\mathsf{T}} F^{-1} b}}$ of Section 1 is about 2.9×10^3 for every ω, with finite KL of about 8.6×10^3 nats. The KL boundaries T_ω instead lie between 6.04 and 6.18. This step misses them because the rare state enters the initial curvature $F_\omega(0) < 26 p_{\mathrm{M}}$ only through its probability, whereas its tilted weight $\omega p_{\mathrm{M}} e^{3t}$ grows exponentially until KL rises sharply near the budget. Appendix C shows the same failure for a binary tilt.

3.2 Complexity of computing useful steps at a fixed KL budget

The four-state result concerns draws. We next ask whether a useful step is also hard to compute when sampling is easy and the direction is exact. The simplest such model tilts the base law toward one event and maximizes the event's probability. Its natural gradient is identically 1, so the direction is known, while the useful step lengths depend on the unknown event mass. Theorem 3.3 shows that calibration again costs as much as estimating this mass, and Corollary 3.4 encodes satisfiability in it.

Theorem 3.3 (Sample complexity of useful event-tilt selection)**.** *Set $p_0 = 2^{-10}$. Let P_0 be a finite law and A a known event with unknown mass $r = P_0(A) \in (0, p_0]$. For $t \ge 0$, tilt the law to $P_t(s) = e^{t\mathbf{1}_A(s)} P_0(s)/(1 - r + re^t)$ and maximize $J(t) = P_t(A)$. Define the event probability, KL cost, gain and optimal feasible gain by*

$$\begin{aligned} q_r(t) &= \frac{re^t}{1 - r + re^t}, \quad \mathcal{K}_r(t) = \mathrm{KL}(P_0 \| P_t) = \log(1 - r + re^t) - rt, \\ \Delta_r(t) &= q_r(t) - r, \qquad\qquad \Delta_r^\star = \max_{t \ge 0:\, \mathcal{K}_r(t) \le 1/10} \Delta_r(t). \end{aligned} \tag{3}$$

A procedure receives the event, the tilted family and the exact natural gradient 1, and otherwise only draws from P_0. It returns a scalar step $t \ge 0$. For $0 < \varepsilon \le 1/4$, this step is useful at mass r when $\mathcal{K}_r(t) \le 1/10$ and $\Delta_r(t) \ge (1 - \varepsilon)\Delta_r^\star$. Fix $0 < p \le p_0/2$ and $0 < \delta \le 1/3$.

(a) Lower bound. *Consider Bernoulli base laws with masses p and $(1 + 4\varepsilon)p$ and event $A = \{1\}$. A procedure takes at most m draws. If it returns a useful step with probability at least $1 - \delta$ under both laws, then $m = \Omega(\log(1/\delta)/(p\varepsilon^2))$.*

(b) Upper bound. *Let $\widehat{p}$ be the event frequency in $m \ge 1$ draws. Use Bernoulli-KL inversion (Garivier & Cappé, 2011) to form $\mathcal{I}_m = \{r \in [0,1] : m d_{\mathrm{B}}(\widehat{p} \| r) \le \log(2/\delta)\}$, where $d_{\mathrm{B}}(a\|b) = a\log(a/b) + (1-a)\log((1-a)/(1-b))$ with boundary values defined by limits. Choose the largest $t \ge 0$ such that $\mathcal{K}_r(t) \le 1/10$ for every $r \in \mathcal{I}_m$, and round downward by at most $\varepsilon/32$ to a nonnegative step. For every finite base law with event mass p, this step is useful with probability at least $1 - \delta$ whenever m exceeds a universal constant multiple of $\log(1/\delta)/(p\varepsilon^2)$. The constants in both bounds are uniform over the stated parameter ranges.*

Since $p \le p_0/2$, the condition $\varepsilon \le 1/4$ keeps the hard pair p and $(1 + 4\varepsilon)p$ inside $(0, p_0]$, while $p_0 = 2^{-10}$ and the finite-precision rounding by $\varepsilon/32$ only make the constants explicit (Appendix B.3). As $p \downarrow 0$, the optimal gain at any fixed budget $\eta > 0$ in place of $1/10$ tends to $1 - e^{-\eta}$. By the KL chain rule, restricting updates to event tilts preserves the best feasible event-probability gain over all laws on the base support (Appendix B.2). Supplying the initial gradient or scalar Fisher information would reveal the event mass, because both equal $p(1-p)$, which increases strictly on the stated range.

Proof sketch. Disjoint useful-step intervals reduce the lower bound to Bernoulli testing. For the upper bound, the worst-case KL at $t > 0$ is attained at $r = 1/t - 1/(e^t - 1)$, clipped to $\mathcal{I}_m$. Doubling and bisection locate the $1/10$ boundary. Coverage ensures feasibility, while concentration and boundary sensitivity bound the gain lost to conservatism and rounding. Appendix B.3 gives the proof, and Theorem B.2 shows that even three quarters of the optimal gain requires $\Omega(1/p)$ draws. □

From event rarity to computational hardness. To turn this sample cost into computational hardness, we map a Boolean formula φ in polynomial time to an event predicate whose mass under the uniform law encodes $\#\mathrm{SAT}(\varphi)$. An always-accepted anchor state keeps the mass positive, and satisfiability at least doubles it. Eleven tag bits keep the mass below p_0 (Appendix A), and the gain target $3/4$ is the most lenient one Theorem 3.3 covers. Padding with unused variables preserves satisfiability, so requiring $n \geq 3$ loses no generality.

Corollary 3.4 (Computational hardness of useful event-tilt selection)**.** *Let φ be a Boolean formula on $n \geq 3$ variables and let P_0 be uniform on $\{0,1\}^{11} \times \{0,1\}^n$. Define the event*

$$A_\varphi = \{(b, x) : b = 0^{11},\ \varphi(x) = 1\} \cup \{(1^{11}, 0^n)\}.$$

With $D_{\mathrm{B}} = 2^{n+11}$, its probability is $p_\varphi = (1 + \#\mathrm{SAT}(\varphi))/D_{\mathrm{B}} < 2^{-10}$. For $t \geq 0$, use the tilt $P_{\varphi,t}(s) \propto e^{t\mathbf{1}_{A_\varphi}(s)} P_0(s)$ and objective $J_\varphi(t) = P_{\varphi,t}(A_\varphi)$. Its natural gradient equals 1, and exact base sampling and event membership are polynomial-time computable from $\langle\varphi\rangle$.

The input is the formula encoding $\langle\varphi\rangle$ and the exact natural gradient. The required output is a nonnegative rational step $\widehat{t}$ whose bit length is polynomial in the input length. It must satisfy $\mathcal{K}_{p_\varphi}(\widehat{t}) \leq 1/10$ and $\Delta_{p_\varphi}(\widehat{t}) \geq (3/4)\Delta^\star_{p_\varphi}$, with KL cost and gain as in Equation (3). This problem is NP-hard. Unless $\mathrm{NP} \subseteq \mathrm{BPP}$ (Arora & Barak, 2009), no polynomial-time procedure with these inputs and draws returns such a step with probability at least $2/3$ on every instance.

Proof sketch. We run a hypothetical calibration procedure on the model encoded by φ and compare its returned step with the threshold $t_{\mathrm{sep,B}} = (n+7)\log 2$, which is computable from n. Useful outputs at the unsatisfiable mass $1/D_{\mathrm{B}}$ exceed $t_{\mathrm{sep,B}} + 0.03$, whereas, by concavity of the KL cost in the event mass, feasible outputs at every satisfiable mass $p_\varphi \geq 2/D_{\mathrm{B}}$ lie below $t_{\mathrm{sep,B}} - 0.0079$. A polynomial-time calibration procedure would therefore decide SAT. A rational threshold accurate to 10^{-3} suffices, and each base draw uses $n + 11$ fair bits. Appendix B.2 proves the finite-precision separation and efficient simulation of exact draws. □

3.3 Complexity of natural-gradient direction recovery

The preceding results take the direction as given. We now ask whether the direction itself is easy to recover when the Fisher matrix is well conditioned. A bounded condition number controls only the ratio of eigenvalues, and rare informative observations can still make both eigenvalues small. In the construction below, the covariate $(1,1)^\mathsf{T}$ couples the two Fisher coordinates, so the unit natural gradient turns as the number of satisfying assignments grows and thereby encodes satisfiability. Here $D_{\mathrm{L}} = 2^{n+5}$ counts the n formula bits, three tag bits and two branch bits (Appendix A).

Theorem 3.5 (Direction recovery with a well-conditioned Fisher matrix)**.** *Fix a Boolean formula φ on $n \geq 3$ variables. Write $K = 1 + \#\mathrm{SAT}(\varphi)$ and $D_{\mathrm{L}} = 2^{n+5}$. With the standard basis e_1, e_2 of $\mathbb{R}^2$, let X have law Q_φ on $\{0, e_1, e_2, (1,1)^\mathsf{T}\}$ given by*

$$\begin{aligned} Q_\varphi(e_1) = Q_\varphi(e_2) = \frac{K}{D_{\mathrm{L}}}, \qquad Q_\varphi((1,1)^\mathsf{T}) = \frac{1}{D_{\mathrm{L}}}, \\ Q_\varphi(0) = 1 - \frac{2K+1}{D_{\mathrm{L}}}. \end{aligned} \tag{4}$$

Consider the binary logistic model $\pi_\theta(Y = 1 \mid X) = \mathrm{sigm}(\theta^\mathsf{T} X)$, where $\theta \in \mathbb{R}^2$ and $\mathrm{sigm}(t) = (1 + e^{-t})^{-1}$, with query objective $\ell(\theta) = \log \mathrm{sigm}(\theta_1)$. Let $F^{\mathrm{L}}_\varphi(\theta)$ be the population Fisher information matrix, averaging over $X \sim Q_\varphi$ and $Y \sim \pi_\theta(\cdot \mid X)$. At $\theta = 0$, the exact gradient $b = \nabla\ell(0) = e_1/2$ is supplied in both recovery problems below. Define $v^{\mathrm{L}}_\varphi = F^{\mathrm{L}}_\varphi(0)^{-1} b$ and the unit natural gradient $\nu^{\mathrm{L}}_\varphi = v^{\mathrm{L}}_\varphi / \|v^{\mathrm{L}}_\varphi\|_2$. Given φ, draws at zero, with law $Q_\varphi \otimes \mathrm{Bern}(1/2)$, are polynomial-time computable.

(a) Fisher matrix and direction. *For every such formula,*

$$F^{\mathrm{L}}_\varphi(0) = \frac{1}{4D_{\mathrm{L}}} \begin{pmatrix} K+1 & 1 \\ 1 & K+1 \end{pmatrix}, \qquad \nu^{\mathrm{L}}_\varphi = \frac{(K+1, -1)^\mathsf{T}}{\sqrt{(K+1)^2 + 1}}. \tag{5}$$

The spectral condition number satisfies $\kappa_2(F^{\mathrm{L}}_\varphi(0)) = 1 + 2/K \leq 3$, and $\lambda_{\min}(F^{\mathrm{L}}_\varphi(0)) = \Theta(2^{-n})$ for $K = 1, 2$.

(b) Direction recovery from the description. *Consider the problem whose input consists of* $(1^n, \langle\varphi\rangle)$ *and* b*, where* 1^n *encodes* n *in unary. It asks for a polynomial-length rational vector* $\widehat{\nu}$ *with* $\|\widehat{\nu} - \nu_\varphi^{\mathrm{L}}\|_2 \le 1/32$*. This problem is NP-hard. Unless* NP $\subseteq$ BPP *(Arora & Barak, 2009), no polynomial-time procedure achieves this accuracy with probability at least* $2/3$ *on every input, even with draws at* $\theta = 0$*.*

(c) Sample-only direction recovery. *Consider the two laws with* $K = 1$ *and* $K = 2$*. A procedure receives* n*, this family and* b*, and learns which law holds only from draws* (X, Y) *at* $\theta = 0$*. Any procedure that takes at most* m *draws and returns* $\widehat{\nu}$ *with* $\|\widehat{\nu} - \nu_\varphi^{\mathrm{L}}\|_2 \le 1/32$ *with probability at least* $2/3$ *under each law must have* $m \ge D_{\mathrm{L}}/6$*.*

(d) Coordinate-threshold decision problem. *The language of valid encodings* $(1^n, \langle\varphi\rangle)$*, with* $n \ge 3$ *and* φ *on* n *variables, for which* $(\nu_\varphi^{\mathrm{L}})_2 > -3/8$ *consists exactly of the satisfiable instances and is NP-complete under polynomial-time many-one reductions.*

Proof sketch. Given φ, we build its model, run a recovery procedure and threshold the output's second coordinate at $-3/8$. By Equation (5), this coordinate is $-1/\sqrt{5}$ when $K = 1$ and at least $-1/\sqrt{10}$ when $K \ge 2$. Both values lie more than $1/32$ from $-3/8$, so an accurate output reveals satisfiability. Satisfying assignments certify membership in the threshold language. From samples, the $K = 1, 2$ laws differ in total variation by $2/D_{\mathrm{L}}$ per draw, so testing needs $m \ge D_{\mathrm{L}}/6$. Appendix D gives the proof, and Figure A1 illustrates the threshold. □

The hardness extends to finite updates. At KL budget $1/(100D_{\mathrm{L}})$ and with no supplied direction, returning a feasible update within $1/200$ of the constrained query optimum is NP-hard, because the small budget keeps every feasible first coordinate at most $1/5$ when $K \ge 2$, whereas such an update for $K = 1$ needs a larger one. Succeeding from draws alone with probability $2/3$ needs $m \ge D_{\mathrm{L}}/6$ (Corollaries D.1 and D.2).

3.4 Damped estimation and KL certification

The remaining results are positive. The logistic construction makes direction recovery hard through small Fisher eigenvalues. Damping is the standard remedy (Martens, 2020; Martens & Grosse, 2015). Replacing F with $F + \lambda I$ for a damping parameter $\lambda > 0$ raises every eigenvalue to at least λ, and we ask how many draws then suffice to estimate the damped direction. The answer is tight up to a logarithmic factor, and Appendix E bounds the damping bias relative to the undamped solution.

Proposition 3.6 (Damped natural gradient estimation)**.** *Let* P *be an unknown probability law on* $\mathcal{X}$ *and* $h : \mathcal{X} \to [0,1]^d$ *a known measurable statistic, with integer* $d \ge 1$*. For known* $\lambda > 0$*, set* $F = \mathrm{Cov}_P(h)$ *and* $\mathsf{A} = F + \lambda I_d$*. Fix* $0 < \varepsilon \le 1/16$ *and* $0 < \delta \le 1/3$*. From* $m = 2s$ *draws* $S_1, \dots, S_{2s} \sim P$*, where* $s \ge 1$ *is an integer, form*

$$\xi_i = \frac{h(S_{2i}) - h(S_{2i-1})}{\sqrt{2}}, \qquad \widehat{F} = \frac{1}{s}\sum_{i=1}^{s} \xi_i \xi_i^{\mathsf{T}}, \qquad \widehat{v}(b) = (\widehat{F} + \lambda I_d)^{-1} b.$$

For a universal constant C*, if* $m \ge C(1 + d/\lambda)\varepsilon^{-2}\log(2d/\delta)$*, then with probability at least* $1 - \delta$*,*

$$\|\widehat{v}(b) - \mathsf{A}^{-1}b\|_{\mathsf{A}} \le \varepsilon \|\mathsf{A}^{-1}b\|_{\mathsf{A}} \qquad \text{for every } b \in \mathbb{R}^d \setminus \{0\}, \tag{6}$$

where $\|z\|_{\mathsf{A}}^2 = z^{\mathsf{T}}\mathsf{A}z$ *and* b *is the supplied objective gradient.*

Conversely, for each $0 < \lambda \le d/8$*, take* $h(z) = z\mathbf{1}_d$ *for* $z \in \{0, 1\}$ *and* $b = \mathbf{1}_d/\sqrt{d}$*, where* $\mathbf{1}_d$ *is the all-ones vector. Consider a procedure that learns an unknown Bernoulli law* P *only from at most* m *draws. If it satisfies Equation (6) for this fixed* b *with probability at least* $2/3$ *for every such* P*, then* $m = \Omega(d/(\lambda\varepsilon^2))$*.*

Proof sketch. Paired differences give an unbiased covariance estimate, and matrix concentration bounds the solve error (Hsu et al., 2014; Tropp, 2012). A Bernoulli statistic repeated across coordinates gives the lower bound, and the ranges $\varepsilon \le 1/16$ and $\lambda \le d/8$ keep that hard pair's masses small and on the damping scale. Appendix E.1 proves both bounds. □

Certifying the KL budget. The lower bounds rule out cheap useful steps in the worst case, yet respecting the budget can still be certified from draws. For a softmax classifier with affine logits, a known bound on the feature norm limits how far each candidate step can move the logits, and hence its KL at every input, before any sampling. Fix N_{cand} candidate steps along a supplied direction, and draw m calibration inputs independently of the classifier, direction and candidates. Empirical Bernstein bounds (Maurer & Pontil, 2009), each at failure probability δ/N_{cand}, then bound the population KL of every candidate at once. Their slack grows at most logarithmically in N_{cand}/δ and shrinks as $m^{-1/2}$, or as m^{-1} when the sample variance is small. Any candidate whose bound is at most η, even one chosen using the same draws, is feasible with probability at least $1-\delta$, and the zero step is always accepted (Proposition F.1). The certificate controls feasibility, not gain.

4 Empirical Illustrations

The experiments examine three distinctions in the theory: calibration sample cost, direction accuracy versus finite gain, and feasibility under sampled KL. Each trial computes the KL cost and gain of its returned update u, and the gain benchmark, exactly under the reference law. The trial is a *joint success* if u is both feasible and within the gain tolerance, that is, useful (Definition 3.1), and its *feasible gain* $\Delta(u)\,\mathbf{1}\{\mathcal{K}(u) \le \eta\}$ is zero for infeasible updates.

Calibration sample scales. The upper-bound rules of Theorems 3.2 and 3.3 stop conservatively short of an estimated KL boundary, since stopping at the estimate itself overshoots about half the time. Their rates also predict similar success across rarity levels at a fixed normalized budget, $mp_{\mathrm{M}}\varepsilon^2$ or $mp\varepsilon^2$ with m draws. Both four-state rules receive exact local information and estimate the mass of $h = 3$ from the same count. The plug-in rule stops at the estimated KL boundary, whereas the conservative rule adds an upward offset to the mass estimate and rounds the step downward. At $\omega = 1.25$, joint success reaches 100% for the conservative rule at the largest sample budget, while plug-in success remains near 50% (Table A1(a)). At that budget, the plug-in rule's mean positive KL excess is only about 0.004% of the budget (Table A3), so its failures are strict but tiny boundary crossings. For binary laws, the confidence rule receives the event count and sample size but not the true mass. At $mp\varepsilon^2 = 16$, joint success ranges from 95.8% to 97.1% as p varies from 10^{-6} to 2^{-11} (Table A1(b) and Appendix H).

Direction accuracy and finite gain. Theorem 3.5 concerns direction accuracy, whereas usefulness concerns the finite step. We separate the two in the logistic query of Section 3.3, with the globally optimal KL-feasible query gain as benchmark. Fix $K = 1$, $D_{\mathrm{L}} = 65{,}536$ and $\eta = 0.05/D_{\mathrm{L}}$. At $m = 16D_{\mathrm{L}}$, only 31.9% of estimated pseudoinverse directions (Figure 1(c)) meet the theorem's Euclidean tolerance $1/32$. Population-KL calibration nevertheless attains 99.5% of optimal gain on average and meets the 90% gain target in every trial. The mean cosine between the estimated and population rays in the Fisher geometry is 99.46%, which explains why the Euclidean pass rate need not predict finite gain (Appendix I.1 and Table A4). Even along the exact direction, confidence calibration is useful in only 1.1% of trials at $m = 16D_{\mathrm{L}}$, rising to 98.9% at $m = 256D_{\mathrm{L}}$ (Table A5). It uses counts even though this direction identifies the cell masses, so its failures measure its conservatism rather than direction error, and Figure 1(c) shows the same pattern with estimated directions. Thus, at $m = 16D_{\mathrm{L}}$, calibrating the length, not estimating the direction, limits usefulness.

Calibration in learned classifier heads. Learned classifier heads test whether stopping at a sampled KL boundary is reliable without constructed rarity. We freeze ResNet-18 representations (He et al., 2016) on CIFAR-10 (Krizhevsky, 2009) and DistilBERT representations (Sanh et al., 2019) on SST-2 (Wang et al., 2018; Socher et al., 2013). Each trial restores the original head and increases an initially misclassified query's true-label log probability. Directions computed on the full training set isolate calibration, and the benchmark is the optimal feasible gain along the supplied ray. The empirical rules share calibration inputs and target the full KL budget or reserve a 10% margin. Under exact KL, convexity of the cost and concavity of the query gain ensure that this margin retains at least 90% of the full-budget ray optimum (Lemma J.1). At the primary budget $\eta = 0.01$ and the largest sample size $m = 16{,}384$, the margin rule's joint success is 99.53% on CIFAR-10 and 93.75% on SST-2, compared with 48.59% and 50.94% at the full budget, which stays near 50% at every sample size because the sampled boundary lies beyond the population one about half the time (Figure 2 and Table A10). Joint success uses a 75% gain target, and every margin update at this setting also meets a stricter 90% target, so the remaining failures are KL violations. Including infeasible updates,

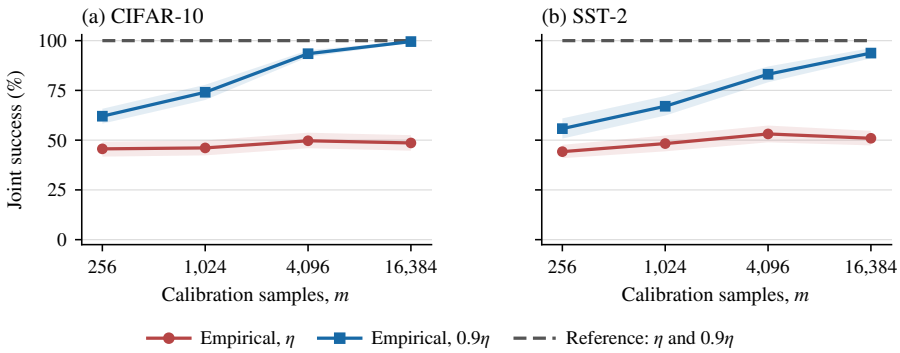

Figure 2: **Calibration along supplied directions in classifier heads. (a)** CIFAR-10 with ResNet-18 features and **(b)** SST-2 with DistilBERT features. Directions and population KL use the full training set, and calibration uses m sampled inputs with KL budget $\eta = 0.01$ or the margin target 0.9η. Bands are 95% bootstrap intervals over 32 queries, with 20 repetitions each. Joint success requires KL feasibility and at least 75% of the optimal feasible gain along the supplied direction at the full budget. Dashed curves use population KL, and colored curves use sampled KL.

the margin keeps the mean raw-gain ratio to the full-budget optimum above 99% in both models, so it costs almost no gain (magnitudes, other budgets and query uncertainty in Appendix J).

5 Discussion

Our results locate a cost that direction-centered analyses leave out. Choosing a useful length needs information about rare states that local quantities do not carry, acquiring it costs as many draws as estimating a rare probability, and computing a useful step can be *intractable* even when sampling is easy and efficient. A natural-gradient update therefore has two distinct costs, estimating the direction and calibrating its length, and the second can be large even when the first is zero.

Any procedure that guarantees useful steps over a larger class containing the four-state family must succeed on the family, so the sample lower bound transfers to that class under the same access and permitted updates. The rate $\log(1/\delta)/(p\varepsilon^2)$ is classical for rare-probability estimation, and our contribution is that calibration inherits it, up to constants, even with every local quantity supplied. The hardness results hold with efficient exact sampling, so they do not restate the hardness of inference (Section 2). Classifier heads cross the boundary about half the time without constructed rarity.

For natural policy gradient, TRPO and K-FAC (Kakade, 2001; Schulman et al., 2015; Martens & Grosse, 2015), the length deserves the scrutiny now given to the direction. Sample budgets should include the draws and computation spent on selecting it, and evaluations should state the KL budget, gain benchmark and success probability together. Violation frequencies need their magnitudes, because under the conditions in Appendix J.2 the probability of crossing the boundary tends to one half while the excess KL vanishes. A small KL margin is a cheap safeguard (Section 4), and independent calibration draws can certify the budget (Section 3.4).

Limitations. Our bounds are worst case, concern single updates and hold under the stated access and output restrictions. Probability tables or extra evaluation oracles can change the information cost, and event-tilt hardness concerns returning a tilt parameter. Under the reverse KL orientation, the optimal event-tilt gain vanishes as $p \downarrow 0$ (Appendix B.2). The experiments fix checkpoints, representations and queries under a finite reference law, and do not test generalization or training trajectories.

Open questions. We highlight three questions for future research. First, how do single-update bounds accumulate over a training run, under either KL orientation? Second, what does it cost to select the direction and the length from the same draws, when confidence bounds must remain valid under reuse (Howard et al., 2021; Dwork et al., 2015)? Third, can importance sampling that oversamples the rare states reduce the draws needed for a useful step (Glynn & Iglehart, 1989; Chatterjee & Diaconis, 2018)?

Ethics statement

This theoretical work does not involve human-participant studies or the collection of new personal data. The accompanying experiments use public benchmark datasets and pretrained models. We do not foresee direct ethical risks specific to the theoretical contributions, although downstream applications should be assessed within their intended deployment contexts.

AI use statement

We used a large language model (LLM) to revise prose and improve the TikZ and graphics visualization code. We did not use it to develop theoretical models or write proofs. The authors reviewed all AI-assisted material and take full responsibility for the manuscript and its artifacts.

Reproducibility statement

The appendix contains complete proofs and experimental protocols. Appendix G gives hardware, software, pinned model and data revisions, the seeding scheme and numerical tolerances. Protocols and parameter grids appear in Appendices H to J. Code, dependency specifications and detailed metadata will be released shortly.

References

Alekh Agarwal, Sham M. Kakade, Jason D. Lee, and Gaurav Mahajan. On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *Journal of Machine Learning Research*, 22(98):1–76, 2021. URL https://jmlr.org/papers/v22/19-736.html.

Shun-ichi Amari. Differential Geometry of Curved Exponential Families—Curvatures and Information Loss. *The Annals of Statistics*, 10(2):357–385, 1982. doi: 10.1214/aos/1176345779.

Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998. doi: 10.1162/089976698300017746.

Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009. doi: 10.1017/CBO9780511804090.

Jose Blanchet, Coralia Cartis, Matt Menickelly, and Katya Scheinberg. Convergence rate analysis of a stochastic trust-region method via supermartingales. *INFORMS Journal on Optimization*, 1(2): 92–119, 2019. doi: 10.1287/ijoo.2019.0016.

Yuheng Bu, Shaofeng Zou, Yingbin Liang, and Venugopal V. Veeravalli. Estimation of KL divergence: Optimal minimax rate. *IEEE Transactions on Information Theory*, 64(4):2648–2674, 2018. doi: 10.1109/TIT.2018.2805844.

Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi. Fast global convergence of natural policy gradient methods with entropy regularization. *Operations Research*, 70(4): 2563–2578, 2022. doi: 10.1287/opre.2021.2151.

Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018. doi: 10.1214/17-AAP1326.

Ruobing Chen, Matt Menickelly, and Katya Scheinberg. Stochastic optimization using a trust-region method and random models. *Mathematical Programming*, 169(2):447–487, 2018. doi: 10.1007/s10107-017-1141-8.

Gregory F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2–3):393–405, 1990. doi: 10.1016/0004-3702(90)90060-D.

Paul Dagum and Michael Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60(1):141–153, 1993. doi: 10.1016/0004-3702(93)90036-B.

Paul Dagum, Richard Karp, Michael Luby, and Sheldon Ross. An optimal algorithm for Monte Carlo estimation. *SIAM Journal on Computing*, 29(5):1484–1496, 2000. doi: 10.1137/S0097539797315306.

Adnan Darwiche. A differential approach to inference in Bayesian networks. *Journal of the ACM*, 50 (3):280–305, 2003. doi: 10.1145/765568.765570.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. doi: 10.1017/CBO9780511802843.

Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth. Preserving statistical validity in adaptive data analysis. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pp. 117–126, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335362. doi: 10.1145/2746539.2746580.

Aurélien Garivier and Olivier Cappé. The KL-UCB algorithm for bounded stochastic bandits and beyond. In Sham M. Kakade and Ulrike von Luxburg (eds.), *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *Proceedings of Machine Learning Research*, pp. 359–376, Budapest, Hungary, 09–11 Jun 2011. PMLR. URL https://proceedings.mlr.press/v19/garivier11a.html.

Peter W. Glynn and Donald L. Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35(11):1367–1392, 1989. doi: 10.1287/mnsc.35.11.1367.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021. doi: 10.1214/20-AOS1991.

Daniel Hsu, Sham M. Kakade, and Tong Zhang. Random design analysis of ridge regression. *Foundations of Computational Mathematics*, 14(3):569–600, 2014. doi: 10.1007/s10208-014-9192-1.

Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986. doi: 10.1016/0304-3975(86)90174-X.

Billy Jin, Katya Scheinberg, and Miaolan Xie. Sample complexity analysis for adaptive optimization algorithms with stochastic oracles. *Mathematical Programming*, 209(1–2):651–679, 2025. doi: 10.1007/s10107-024-02078-z.

Sham Kakade. A natural policy gradient. In *Proceedings of the 15th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, pp. 1531–1538, Cambridge, MA, USA, 2001. MIT Press. URL https://dl.acm.org/doi/10.5555/2980539.2980738.

Sham Kakade and John Langford. Approximately optimal approximate reinforcement learning. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pp. 267–274. Morgan Kaufmann, 2002. URL https://dl.acm.org/doi/10.5555/645531.656005.

P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P. Grünwald. On predictive distributions and Bayesian networks. *Statistics and Computing*, 10(1):39–54, 2000. doi: 10.1023/A:1008984400380.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://cave.cs.toronto.edu/kriz/learning-features-2009-TR.pdf.

Frederik Kunstner, Lukas Balles, and Philipp Hennig. Limitations of the empirical Fisher approximation for natural gradient descent. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc. URL https://dl.acm.org/doi/10.5555/3454287.3454661.

Maren Mahsereci and Philipp Hennig. Probabilistic line searches for stochastic optimization. In *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pp. 181–189, Cambridge, MA, USA, 2015. MIT Press. URL https://dl.acm.org/doi/10.5555/2969239.2969260.

James Martens. New insights and perspectives on the natural gradient method. *Journal of Machine Learning Research*, 21(146):1–76, 2020. URL https://jmlr.org/papers/v21/17-678.html.

James Martens and Roger Grosse. Optimizing neural networks with Kronecker-factored approximate curvature. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2408–2417, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/martens15.html.

Andreas Maurer and Massimiliano Pontil. Empirical Bernstein bounds and sample variance penalization. In *Annual Conference on Learning Theory*, 2009. URL https://www.learningtheory.org/colt2009/papers/012.pdf.

Courtney Paquette and Katya Scheinberg. A stochastic line search method with expected complexity analysis. *SIAM Journal on Optimization*, 30(1):349–376, 2020. doi: 10.1137/18M1216250.

Matteo Pirotta, Marcello Restelli, and Luca Bascetta. Adaptive step-size for policy gradient methods. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/f64eac11f2cd8f0efa196f8ad173178e-Paper.pdf.

Dan Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1–2):273–302, 1996. doi: 10.1016/0004-3702(94)00092-1.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019. doi: 10.48550/arXiv.1910.01108.

John Schulman, Sergey Levine, Philipp Moritz, Michael Jordan, and Pieter Abbeel. Trust region policy optimization. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1889–1897, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/schulman15.html.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In David Yarowsky, Timothy Baldwin, Anna Korhonen, Karen Livescu, and Steven Bethard (eds.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642, Seattle, Washington, USA, October 2013. Association for Computational Linguistics. doi: 10.18653/v1/D13-1170.

Philip Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. High confidence policy improvement. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2380–2388, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/thomas15.html.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. doi: 10.1007/s10208-011-9099-z.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. doi: 10.1007/b13794.

Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, 1979. doi: 10.1137/0208032.

Tim van Erven and Peter Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014. doi: 10.1109/TIT.2014.2320500.

Sharan Vaswani, Aaron Mishkin, Issam Laradji, Mark Schmidt, Gauthier Gidel, and Simon Lacoste-Julien. Painless stochastic gradient: Interpolation, line-search, and convergence rates. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/2557911c1bf75c2b643afb4ecbfc8ec2-Abstract.html.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008. doi: 10.1561/2200000001.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In Tal Linzen, Grzegorz Chrupała, and Afra Alishahi (eds.), *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 353–355, Brussels, Belgium, November 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-5446.

Edwin B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927. doi: 10.1080/01621459.1927.10502953.

APPENDIX

The appendices follow the order of the main text. They first prove the theoretical results and then report reproducibility details and the experiments that illustrate them. Appendix A fixes the model of a procedure, the padding lemma used by every sample lower bound and the satisfiability encoding used by both hardness results. Appendix B treats calibration with an exact natural gradient at a fixed budget. Within it, Appendix B.1 proves Theorem 3.2 through Theorem B.1(b)–(c). Appendix B.2 proves Corollary 3.4 using the constant-fraction separation in Theorem B.2, and Appendix B.3 proves Theorem 3.3. Appendix C compares quadratic and finite-step KL. Appendix D proves Theorem 3.5 on direction recovery and extends it to useful finite updates (Corollaries D.1 and D.2). Appendix E proves Proposition 3.6 through Proposition E.1 and Corollary E.3 and bounds the bias introduced by damping. Appendix F states the population-KL certificate summarized in Section 3.4 (Proposition F.1) and proves it through Proposition F.2. Appendix G gives reproducibility details. The experiments then cover count simulations (Appendix H), logistic controls and damping simulations (Appendix I) and learned classifier heads (Appendix J).

Notation across the appendices. The sample count m counts individual observations. Thus s independent pairs contain $2s$ observations. The four-state family uses supplied scale p_{M} and unknown coefficient ω, while p denotes an event probability and c a formula code. The count K includes one always-accepted state in addition to the satisfying assignments. KL costs are written $\mathcal{K}$, and $D_{\mathrm{L}}, D_{\mathrm{B}}$ denote the logistic and event-tilt encoding sizes. Absolute KL budgets are η, and logistic experiments also use $\bar{\eta} = D_{\mathrm{L}}\eta$.

A PROCEDURES AND LOWER-BOUND TOOLS

Every sample lower bound in the appendices uses the following model of a procedure, which makes the description in Section 2 precise. Draws $X_1, X_2, \dots$ are independent with a common law P on a measurable space, and an independent seed U has a law that does not depend on P. Write $\mathcal{F}_k = \sigma(U, X_1, \dots, X_k)$ for the information available after k draws. A procedure with sample budget m stops after $N \le m$ draws, where N is a stopping time for $(\mathcal{F}_k)_{k\ge 0}$. Thus whether the procedure takes another draw depends only on the seed and the draws seen so far. Its output lies in a Euclidean space and is $\mathcal{F}_N$-measurable. Supplied quantities, such as an exact natural gradient, are identical under the laws being compared and enter as fixed inputs. Write $\mathcal{L}_P(Y)$ for the law of a random variable Y when the draws have law P.

Lemma A.1 (Padding and two-point testing)**.** *The output equals $g(U, X_1, \dots, X_m)$ for a measurable map g that does not depend on P. Consequently, for two laws P and P' of the draws, the output Y satisfies*

$$\begin{aligned} &\mathrm{KL}(\mathcal{L}_P(Y) \| \mathcal{L}_{P'}(Y)) \le m\,\mathrm{KL}(P\|P'), \\ &\mathrm{TV}(\mathcal{L}_P(Y), \mathcal{L}_{P'}(Y)) \le 1 - (1 - \mathrm{TV}(P, P'))^m \le m\,\mathrm{TV}(P, P'). \end{aligned} \tag{A1}$$

If S and S' are disjoint measurable sets of outputs, then $\mathbb{P}_P(Y \notin S) + \mathbb{P}_{P'}(Y \notin S') \ge 1 - \mathrm{TV}(\mathcal{L}_P(Y), \mathcal{L}_{P'}(Y))$.

Proof. On $\{N = k\}$, the output is $\mathcal{F}_k$-measurable, so the Doob–Dynkin lemma gives a measurable map g_k with $Y = g_k(U, X_1, \dots, X_k)$ there. The map $g = \sum_{k\le m} \mathbf{1}\{N = k\} g_k$ ignores every draw after the stopping time, so padding the transcript to m draws leaves the output unchanged. The seed has the same law under both hypotheses and is independent of the draws, so the chain rule for KL and independence give KL equal to $m\,\mathrm{KL}(P\|P')$ for $(U, X_1, \dots, X_m)$. Data processing (van Erven & Harremoës, 2014) transfers this bound to $Y = g(U, X_1, \dots, X_m)$. For total variation, couple each pair of draws maximally, independently across positions, and use the same seed under both laws. The two transcripts then agree with probability $(1 - \mathrm{TV}(P, P'))^m$, and so do the outputs. Bernoulli's inequality gives the last bound in Equation (A1). Finally, S' lies in the complement of S, so $\mathbb{P}_P(Y \in S) + \mathbb{P}_{P'}(Y \in S') \le 1 + \mathrm{TV}(\mathcal{L}_P(Y), \mathcal{L}_{P'}(Y))$. □

In each lower bound below, useful outputs under two laws form disjoint sets. The lemma then turns a success guarantee under both laws into a bound on the number of draws, whatever the procedure's computational resources.

How the constructions encode satisfiability. Both hardness results start from a classical encoding. An assignment x drawn uniformly from $\{0,1\}^n$ satisfies φ with probability $\#\mathrm{SAT}(\varphi)/2^n$, so fair bits and one formula evaluation sample a law that encodes the number of satisfying assignments. This encoding makes exact and approximate inference in Bayesian networks hard (Cooper, 1990; Dagum & Luby, 1993; Roth, 1996). Computing the count is #P-complete (Valiant, 1979), yet sampling stays easy because draws range over all assignments rather than satisfying ones (Jerrum et al., 1986). We add three ingredients. An *anchor state* is always accepted, so the accepted count $K = 1 + \#\mathrm{SAT}(\varphi)$ is at least one. This keeps event masses positive, as Theorem 3.3 requires, and the logistic Fisher matrix positive definite. Satisfiability then becomes the question whether $K = 1$ or $K \geq 2$. *Tag bits*, drawn uniformly with x, split the states into disjoint blocks with separate acceptance rules, such as one for satisfying assignments and one for the anchor. Each tag bit halves every state's mass. Eleven tag bits are the fewest that keep the total mass of $2^n + 1$ accepted states below $p_0 = 2^{-10}$ in Corollary 3.4. Three tag bits separate the formula block, the anchor and one fixed state in Theorem 3.5. In the logistic model, two *branch bits* choose whether a state produces e_1, e_2, $(1,1)^\mathsf{T}$ or 0, so every covariate probability is a count divided by $D_\mathrm{L} = 2^{n+5}$. The shared covariate $(1,1)^\mathsf{T}$ couples the two Fisher coordinates, so the direction turns as K grows. When there is only one satisfying assignment, a draw reveals it with probability exponentially small in n. Thus sampling cannot cheaply distinguish $K = 1$ from $K \geq 2$ uniformly over formulas. The separating threshold depends only on n, so a polynomial-time procedure that returns a useful step or an accurate direction would decide SAT.

B Fixed-budget calibration with an exact natural gradient

This appendix proves the results of Sections 3.1 and 3.2. Appendix B.1 treats the four-state family, Appendix B.2 gives the event-tilt bounds and the satisfiability reduction, and Appendix B.3 derives the dependence on the required improvement.

B.1 Calibration with exact gradient and scalar Fisher information

We prove Theorem 3.2 by varying the finite-step KL boundary while preserving the moments that determine the supplied gradient and scalar Fisher information. We first derive the local identities and boundary geometry, then state the quantitative bounds in Theorem B.1. Its parts (b)–(c) give the main result on $\omega \in [1, 3/2]$. The wider interval $\omega \leq 2$ also permits the constant-fraction comparison in part (a).

Minimal support for preserving two moments. For fixed distinct statistic values $a_1, \ldots, a_k$, normalization, mean and second moment constrain the probabilities through the matrix with columns $(1, a_j, a_j^2)^\mathsf{T}$. Its Vandermonde rank is $\min\{3, k\}$. For $k \leq 3$, these constraints therefore determine the law uniquely. Four values are the minimum permitting distinct laws with the same first two moments. On $\{0,1,2,3\}$, varying ω in the construction below changes the mass vector in direction $p_\mathrm{M}(-1, 3, -3, 1)$, which lies in this matrix's one-dimensional nullspace. Thus the family realizes this minimum while retaining positive probabilities.

Set $p_{\mathrm{M},\max} = 2^{-30}$. For $0 < p_\mathrm{M} \leq p_{\mathrm{M},\max}$ and $1 \leq \omega \leq 2$, define a law on $h \in \{0,1,2,3\}$ by

$$\big(P_{p_\mathrm{M},\omega}(0), P_{p_\mathrm{M},\omega}(1), P_{p_\mathrm{M},\omega}(2), P_{p_\mathrm{M},\omega}(3)\big) = \big(1 - (5+\omega)p_\mathrm{M}, (3\omega - 2)p_\mathrm{M}, (7 - 3\omega)p_\mathrm{M}, \omega p_\mathrm{M}\big). \tag{A2}$$

The scalar parameter $t \geq 0$ is optimized, while ω indexes the unknown base law. The main text suppresses the fixed scale p_M in P_ω. Tilt by e^{th}, and write

$$\begin{aligned} Z_\omega(t) &= \mathbb{E}_{P_{p_\mathrm{M},\omega}} e^{th}, & P_{p_\mathrm{M},\omega,t}(h) &= e^{th} P_{p_\mathrm{M},\omega}(h)/Z_\omega(t), \\ J_\omega(t) &= \mathbb{E}_{P_{p_\mathrm{M},\omega,t}} h, & \mathcal{K}_\omega(t) &= \mathrm{KL}(P_{p_\mathrm{M},\omega} \| P_{p_\mathrm{M},\omega,t}), \\ \Delta_\omega(t) &= J_\omega(t) - J_\omega(0), & \Delta_\omega^\star &= \max_{t \geq 0:\, \mathcal{K}_\omega(t) \leq 1/10} \Delta_\omega(t). \end{aligned} \tag{A3}$$

This objective is a mean of a four-valued statistic. Let $b_\omega(t) = J'_\omega(t)$ and let $F_\omega(t)$ denote the scalar Fisher information for the coordinate t. A procedure receives p_M, the family specification, the statistic h, and the exact $b_\omega(0), F_\omega(0)$ and natural gradient. Its only further information about ω consists of independent base draws.

Local identities and the feasible boundary. All four masses are positive and sum to one. Direct summation gives $\mathbb{E}h = 12p_\mathrm{M}$ and $\mathbb{E}h^2 = 26p_\mathrm{M}$, independently of ω. The normalizers satisfy

$Z_{\omega'}(t) - Z_\omega(t) = (\omega' - \omega)p_{\mathrm{M}}(e^t - 1)^3$. Thus their difference vanishes through second order at zero while changing the finite-step KL boundary. The tilted score is $h - J_\omega(t)$, so $J'_\omega(t) = F_\omega(t) = \mathrm{Var}_{P_{p_{\mathrm{M}},\omega,t}} h > 0$, giving the shared local quantities in Equation (A4) below. Moreover, $\mathcal{K}_\omega(t) = \log Z_\omega(t) - 12p_{\mathrm{M}}t$ and $\mathcal{K}'_\omega(t) = J_\omega(t) - 12p_{\mathrm{M}} > 0$ for $t > 0$. Since $\mathcal{K}_\omega(t) \geq \log(\omega p_{\mathrm{M}}) + (3 - 12p_{\mathrm{M}})t$, there is a unique positive T_ω with $\mathcal{K}_\omega(T_\omega) = 1/10$. The objective increases, so $\Delta^\star_\omega = \Delta_\omega(T_\omega)$. Thus a useful step must be large enough to achieve the requested gain and small enough to remain below T_ω. For the lower bounds, we show that these allowable intervals are disjoint for selected pairs of laws. For the upper bound, we estimate a boundary that keeps the returned step feasible under the unknown law.

Theorem B.1 (Calibration with matched base gradients and scalar Fisher information)**.** *Fix $0 < p_{\mathrm{M}} \leq p_{\mathrm{M,max}} = 2^{-30}$ and $0 < \delta \leq 1/3$. Use the four-state law Equation (A2), tilt, gain and benchmark in Equation (A3). A step $t \geq 0$ is KL-feasible when $\mathcal{K}_\omega(t) \leq 1/10$. Every law with $\omega \in [1, 2]$ satisfies, for every $t \geq 0$,*

$$b_\omega(0) = F_\omega(0) = 26p_{\mathrm{M}} - 144p_{\mathrm{M}}^2 > 0, \qquad F_\omega(t)^{-1}b_\omega(t) = 1. \tag{A4}$$

Consider total measurable procedures with an independent random seed and at most a deterministic integer m base draws, including early stopping, under the access specified above. Totality means that the procedure returns an output for every sample sequence and every value of its seed.

(a) *If an output $\widehat{t} \geq 0$ is KL-feasible and attains at least $3/4$ of $\Delta^\star_\omega$, with probability at least $1 - \delta$ under each of $\omega = 1, 2$, then*

$$m \geq \frac{\log(1/(2\delta))}{-\log(1 - 4p_{\mathrm{M}})}. \tag{A5}$$

At $\delta = 1/3$, this implies $m \geq 1/(16p_{\mathrm{M}})$.

(b) *For $0 < \varepsilon \leq 1/8$, if an output is KL-feasible and attains at least $(1 - \varepsilon)\Delta^\star_\omega$, with probability at least $1 - \delta$ under both $\omega = 1$ and $\omega = 1 + 4\varepsilon$, then*

$$m \geq \frac{(1 - 2\delta)\log((1 - \delta)/\delta)}{220p_{\mathrm{M}}\varepsilon^2}. \tag{A6}$$

(c) *Conversely, for $0 < \varepsilon \leq 1/8$, a rule using the supplied p_{M} returns a KL-feasible step with gain at least $(1 - \varepsilon)\Delta^\star_\omega$, with probability at least $1 - \delta$ for every $\omega \in [1, 3/2]$, whenever*

$$m \geq \frac{120000\log(2/\delta)}{p_{\mathrm{M}}\varepsilon^2}. \tag{A7}$$

It can return a nonnegative rational step. This guarantee concerns sample count and does not assert a running-time bound.

Proof of Theorem B.1. We first establish a uniform lower bound on gain. We then prove the sample lower bounds by separating useful-step intervals and obtain the upper bound by confidence calibration.

Uniform gain bound. A feasible candidate gives a lower bound on the optimal gain, uniformly in ω. Take $t_{\mathrm{M,cand}} = \frac{1}{3}\log(1/(10\omega p_{\mathrm{M}}))$, and put $z = e^{t_{\mathrm{M,cand}}} \geq 1$. This step gives the state $h = 3$ a fixed unnormalized weight. The weights of $h = 1, 2$, involving lower powers of the tilt, remain small. Then $\omega p_{\mathrm{M}} z^3 = 1/10$, $p_{\mathrm{M}} z \leq p_{\mathrm{M}} z^2$, and $(p_{\mathrm{M}} z^2)^3 = p_{\mathrm{M}}/(100\omega^2) \leq 1/4096^3$. Both middle coefficients in Equation (A2) are at most four, giving

$$Z_\omega(t_{\mathrm{M,cand}}) \leq \frac{11}{10} + \frac{8}{4096} < \frac{221}{200} \leq e^{1/10}. \tag{A8}$$

Thus $t_{\mathrm{M,cand}}$ is feasible. Its first-moment numerator is at least $3/10$, and $Z_\omega(t_{\mathrm{M,cand}}) < 111/100$, so

$$\Delta^\star_\omega \geq \frac{10}{37} - 12p_{\mathrm{M}} \geq \frac{10}{37} - \frac{1}{10000} > \frac{27}{100}. \tag{A9}$$

Part (a): a fixed separator. For part (a), set $t_{\mathrm{sep,M}} = \frac{1}{3}\log(1/(16p_{\mathrm{M}}))$ and $z = e^{t_{\mathrm{sep,M}}}$. This step is too short to attain the required gain under the first law and infeasible under the second. At this step,

$p_{\mathrm{M}} z^3 = 1/16$ and $p_{\mathrm{M}} z \le p_{\mathrm{M}} z^2 \le 1/4096$. Under $\omega = 1$, $Z_1(t_{\mathrm{sep,M}}) \ge 1 - 6p_{\mathrm{M}} \ge 31/32$, and hence

$$J_1(t_{\mathrm{sep,M}}) \le \frac{3/16 + 9/4096}{31/32} < \frac{1}{5} < \frac{3}{4}\frac{27}{100}. \tag{A10}$$

Every useful output under this law therefore exceeds $t_{\mathrm{sep,M}}$. Under $\omega = 2$, $Z_2(t_{\mathrm{sep,M}}) \ge 9/8 - 7p_{\mathrm{M}} > 1.12$. Using $\log(1+x) \ge x - x^2/2$ for $x \ge 0$ gives $\log Z_2(t_{\mathrm{sep,M}}) > 0.11$. Since $p_{\mathrm{M}} \log(1/p_{\mathrm{M}}) \le 2\sqrt{p}_{\mathrm{M}}$,

$$\mathcal{K}_2(t_{\mathrm{sep,M}}) > 0.11 - 8\sqrt{p}_{\mathrm{M}} > 0.1. \tag{A11}$$

The bound on $p_{\mathrm{M}} \log(1/p_{\mathrm{M}})$ follows from $\log x \le x - 1$ at $x = 1/\sqrt{p}_{\mathrm{M}}$. Every feasible output under $\omega = 2$ is below $t_{\mathrm{sep,M}}$. The mass differences between the two base laws are $p_{\mathrm{M}}(-1, 3, -3, 1)$, so their total variation is $4p_{\mathrm{M}}$. Their useful-step sets are disjoint, so Lemma A.1 bounds the error sum below by $(1 - 4p_{\mathrm{M}})^m$, proving Equation (A5). At $\delta = 1/3$, use $\log(3/2) \ge 1/3$, $-\log(1 - 4p_{\mathrm{M}}) \le 4p_{\mathrm{M}}/(1 - 4p_{\mathrm{M}})$, and $p_{\mathrm{M}} \le 1/16$.

Part (b): interval width and boundary motion. For part (b), restrict ω to $[1, 3/2]$. Let $L_{\omega,\varepsilon} \in (0, T_\omega)$ be the unique step with $\Delta_\omega(L_{\omega,\varepsilon}) = (1 - \varepsilon)\Delta^\star_\omega$. The useful-step interval is $[L_{\omega,\varepsilon}, T_\omega]$. The contribution of $h = 3$ to the normalizer and the boundary equation give

$$12p_{\mathrm{M}} T_\omega \le \frac{4p_{\mathrm{M}}}{1 - 4p_{\mathrm{M}}}\left(\frac{1}{10} - \log p_{\mathrm{M}}\right) \le 8\left(\frac{p_{\mathrm{M}}}{10} + 2\sqrt{p}_{\mathrm{M}}\right) < \frac{1}{1024}. \tag{A12}$$

Using $e^x \le (1 - x)^{-1}$ for $0 \le x < 1$, we obtain $Z_\omega(T_\omega) = e^{1/10 + 12p_{\mathrm{M}} T_\omega} \le 9/8$. Since $h \le 3$,

$$J_\omega(T_\omega) \le 3\left(1 - \frac{P_{p_{\mathrm{M}},\omega}(0)}{Z_\omega(T_\omega)}\right) \le \frac{1}{3} + \frac{52}{3} p_{\mathrm{M}} \le \frac{7}{20}. \tag{A13}$$

Write $q_j = P_{p_{\mathrm{M}},\omega,t}(h = j)$ and $z = e^t$. For $0 \le t \le T_\omega$, $\omega p_{\mathrm{M}} z^3 \le Z_\omega(t) \le 9/8$, so $(p_{\mathrm{M}} z^2)^3 \le 81p_{\mathrm{M}}/64 \le 1/512^3$. Also $Z_\omega(t) \ge 1$, $z \ge 1$, $3\omega - 2 \le 5/2$, and $7 - 3\omega \le 4$. Consequently $q_1 + q_2 \le 13/1024$. On the useful-step interval, $189/800 = (7/8)(27/100) \le J_\omega(t) \le 7/20$. Under the tilted law, the identity $\mathbb{E}h^2 = 3J_\omega(t) - 2(q_1 + q_2)$ yields

$$J'_\omega(t) = 3J_\omega(t) - J_\omega(t)^2 - 2(q_1 + q_2) \ge 3\frac{189}{800} - \left(\frac{189}{800}\right)^2 - \frac{13}{512} > \frac{3}{5}. \tag{A14}$$

Here $3J - J^2$ increases on the stated interval. Integrating and using Equation (A13) gives

$$T_\omega - L_{\omega,\varepsilon} \le \frac{\varepsilon \Delta^\star_\omega}{3/5} \le \frac{7}{12}\varepsilon. \tag{A15}$$

The implicit-function theorem applies to $\mathcal{K}_\omega(T_\omega) = 1/10$, since $\partial_t \mathcal{K}_\omega(T_\omega) = \Delta^\star_\omega > 0$. Uniqueness makes the local solutions agree on overlaps, so they define a differentiable boundary function on a neighborhood of $[1, 3/2]$. Differentiating in ω, with $z = e^{T_\omega}$, gives

$$-\partial_\omega T_\omega = \frac{p_{\mathrm{M}}(z - 1)^3}{Z_\omega(T_\omega)(J_\omega(T_\omega) - 12p_{\mathrm{M}})}. \tag{A16}$$

The boundary satisfies $z > 64$: at $t = \log 64$, $Z_\omega(t) \le 1 + 409760p_{\mathrm{M}} < 11/10$, so $\mathcal{K}_\omega(t) < 1/10$. We enlarge the denominator by dropping the negative term $-12p_{\mathrm{M}} Z_\omega(T_\omega)$ and bounding the three remaining coefficients using the stated parameter range. This gives

$$-\partial_\omega T_\omega \ge \frac{(1 - 1/z)^3}{(5/2)/z^2 + 8/z + 9/2} \ge \frac{(63/64)^3}{(5/2)/64^2 + 8/64 + 9/2} > \frac{1}{5}. \tag{A17}$$

Thus $T_1 - T_{1+4\varepsilon} \ge (4/5)\varepsilon$, which exceeds the width in Equation (A15). In particular, $L_{1,\varepsilon} > T_{1+4\varepsilon}$.

Per-observation KL bound. The base mass differences are $\varepsilon p_{\mathrm{M}}(-4, 12, -12, 4)$. Applying $\log x \le x - 1$ termwise gives

$$\begin{aligned} \mathrm{KL}(P_{p_{\mathrm{M}},1} \| P_{p_{\mathrm{M}},1+4\varepsilon}) &\le \varepsilon^2 p_{\mathrm{M}}\left[\frac{16p_{\mathrm{M}}}{1 - (6 + 4\varepsilon)p_{\mathrm{M}}} + \frac{144}{1 + 12\varepsilon} + \frac{144}{4 - 12\varepsilon} + \frac{16}{1 + 4\varepsilon}\right] \\ &\le \varepsilon^2 p_{\mathrm{M}}\left[1 + 144 + \frac{288}{5} + 16\right] < 220 p_{\mathrm{M}} \varepsilon^2. \end{aligned} \tag{A18}$$

Thresholding the output between the disjoint intervals defines a test with both errors at most δ. Lemma A.1 and monotonicity in the two test probabilities give

$$m\,\mathrm{KL}(P_{p_{\mathrm{M}},1}\|P_{p_{\mathrm{M}},1+4\varepsilon}) \geq (1-2\delta)\log((1-\delta)/\delta). \tag{A19}$$

The supplied quantities are identical under both laws. This proves Equation (A6).

Part (c): confidence calibration and rounding. The supplied scale converts the observed frequency of $h=3$ into an estimate of ω. We add a positive offset to this estimate and clip the result to the allowed parameter interval. On the concentration event below, the resulting step stays within the true KL boundary. We then bound the gain lost by this adjustment. For part (c), let $\widehat{q}$ be the observed frequency of $h=3$, and set

$$\omega_U = \min\left\{\frac{3}{2}, \max\left\{1, \frac{\widehat{q}}{p_{\mathrm{M}}} + \frac{\varepsilon}{100}\right\}\right\}. \tag{A20}$$

The true probability of $h=3$ is ωp_{M}. Multiplicative Bernoulli concentration with relative tolerance $\varepsilon/200$ bounds the failure probability by $2\exp(-m\omega p_{\mathrm{M}}\varepsilon^2/120000)$ (Hoeffding, 1963). Under Equation (A7), with probability at least $1-\delta$,

$$\left|\frac{\widehat{q}}{p_{\mathrm{M}}} - \omega\right| \leq \frac{3\varepsilon}{400}, \qquad \omega \leq \omega_U \leq \omega + \frac{7\varepsilon}{400} < \omega + \frac{\varepsilon}{50}. \tag{A21}$$

Return the boundary T_{ω_U} of the upper estimate. Since $\partial_\omega \mathcal{K}_\omega(t) = p_{\mathrm{M}}(e^t-1)^3/Z_\omega(t) \geq 0$, this step is feasible under the true law on the same event. At a boundary, $\partial_\omega \mathcal{K}_\omega \leq p_{\mathrm{M}} e^{3T_\omega}/Z_\omega(T_\omega) = q_3/\omega \leq 1$. Together with Equation (A9), this gives $|\partial_\omega T_\omega| \leq 100/27$, and hence $0 \leq T_\omega - T_{\omega_U} \leq 2\varepsilon/27$. Every law on $[0,3]$ has variance at most $9/4$, as follows by centering at $3/2$. Therefore

$$\Delta^\star_\omega - \Delta_\omega(T_{\omega_U}) \leq \frac{9}{4}(T_\omega - T_{\omega_U}) \leq \frac{\varepsilon}{6} < \varepsilon\Delta^\star_\omega. \tag{A22}$$

For a rational output, take $N_{\mathrm{round}} = \lceil 100/\varepsilon \rceil$ and return $\lfloor N_{\mathrm{round}} T_{\omega_U} \rfloor / N_{\mathrm{round}}$. This is a measurable nonnegative downward rounding with error at most $\varepsilon/100$. It preserves feasibility and adds at most $9\varepsilon/400$ gain loss. The total gain loss is less than $0.19\varepsilon < \varepsilon\Delta^\star_\omega$ by Equation (A9), proving the gain guarantee. □

By Equation (A48), the confidence factor in part (b) is of order $\log(1/\delta)$ uniformly for $0 < \delta \leq 1/3$. Together with part (c), this proves the optimal worst-case sample complexity in Theorem 3.2.

B.2 Calibration bounds and computational reduction

We use event tilts to isolate the cost of scalar calibration when the natural gradient is already known. We give explicit constants for $0 < p \leq p_0 = 2^{-10}$. Theorem B.2 supplies the boundary and confidence bounds used in Theorem 3.3. Its separation argument also supports the proof of Corollary 3.4 below. Let P_0 be a finite base law and A a known event of unknown mass $p = P_0(A) \in (0, p_0]$. For $t \geq 0$, define the event tilt and objective by

$$P_t(s) = \frac{e^{t\mathbf{1}_A(s)}P_0(s)}{1-p+pe^t}, \qquad J(t) = q_p(t) = \frac{pe^t}{1-p+pe^t}. \tag{A23}$$

Its population KL cost, gain and optimal feasible gain are

$$\begin{aligned} \mathcal{K}_p(t) &= \mathrm{KL}(P_0\|P_t) = \log(1-p+pe^t) - pt, \\ \Delta_p(t) &= q_p(t) - p, \qquad \Delta^\star_p = \max_{t\geq 0:\,\mathcal{K}_p(t)\leq 1/10} \Delta_p(t). \end{aligned} \tag{A24}$$

Write $F(t)$ for the scalar Fisher information of this tilt. A procedure receives the family specification, event-membership access and the exact natural gradient 1. Independent base draws are its only further information about p. Neither the initial gradient nor scalar Fisher information is supplied. Write $d_{\mathrm{B}}(a\|b) = a\log(a/b) + (1-a)\log((1-a)/(1-b))$ for Bernoulli KL, with boundary values defined by limits. Write $\mathrm{logit}(q) = \log(q/(1-q))$ for $0 < q < 1$.

The required output is a tilt parameter, which matters even for a binary base law. At KL budget $1/10$, the fixed law $\mathrm{Bern}(0.09)$ is feasible and 75%-useful for every $0 < p \leq 2^{-10}$. Its prescribed tilt, however, is $t = \mathrm{logit}(0.09) - \mathrm{logit}(p)$ and depends on the unknown mass. Thus specifying a useful law need not require the information needed to select its tilt.

Theorem B.2 (Useful-step calibration with an exact natural gradient)**.** *For the event-tilt model above with $0 < p \le p_0 = 2^{-10}$, the following hold under the stated sampling access. A feasible step is a scalar $t \ge 0$ satisfying $\mathcal{K}_p(t) \le 1/10$, and its gain is $\Delta_p(t)$.*

(a) Exact natural gradient and attainable improvement. *The scalar Fisher information satisfies $J'(t) = F(t) = q_p(t)(1 - q_p(t))$, so $F(t)^{-1}J'(t) = 1$. Moreover, $\Delta_p^\star \ge 9/100 - p$ and $\lim_{p\downarrow 0} \Delta_p^\star = 1 - e^{-1/10}$.*

(b) Sample lower bound. *For every $0 < p \le p_0/2$, there exist two base laws with the same known event A and event probabilities p and $2p$. Consider any procedure using at most m draws. Success means returning a feasible step attaining at least $3/4$ of optimal gain. If the procedure succeeds with probability at least $1 - \delta$ under each law, $0 < \delta \le 1/3$, then $m \ge \log(1/(2\delta))/[-\log(1 - p)]$. At $\delta = 1/3$, this gives $m \ge 1/(3p)$, regardless of computational power.*

(c) Confidence-based calibration. *Fix $0 < \delta < 1$. For any known event A, let $\widehat{p}$ be its observed frequency in $m \ge 1$ draws. Use Bernoulli-KL inversion (Garivier & Cappé, 2011) to form the interval $\mathcal{I}_m = \{r \in [0, 1] : m\, d_{\mathrm{B}}(\widehat{p}\|r) \le \log(2/\delta)\}$. Let $\widehat{t}_m$ be the largest $t \ge 0$ satisfying $\mathcal{K}_r(t) \le 1/10$ for every $r \in \mathcal{I}_m$. For every fixed m, this rule is KL-feasible with probability at least $1 - \delta$. If $m \ge (2048/p)\log(8/\delta)$, the returned step is feasible and attains at least $3/4$ of optimal gain with probability at least $1 - \delta$ under P_0.*

Proof of Theorem B.2. The KL boundary locates the optimal gain and separates useful outputs under two event masses. For the upper bound, the returned step is feasible whenever the confidence interval contains the true mass. Controlling the interval's width also gives usefulness on this event.

Boundary geometry. Throughout, $0 < p \le p_0 = 2^{-10}$. The identities $q_p' = q_p(1 - q_p) > 0$, $\mathcal{K}_p' = q_p - p > 0$ for $t > 0$, and $\mathcal{K}_p(t) = (1-p)t + \log p + o(1)$ as $t \to \infty$, together with $\mathcal{K}_p(0) = 0$, give a unique positive boundary $T(p)$ satisfying $\mathcal{K}_p(T(p)) = 1/10$. Strict increase of q_p places the optimum at this boundary. Substituting $t = \log[q(1 - p)/(p(1 - q))]$ gives

$$\mathcal{K}_p(t) = d_{\mathrm{B}}(p\|q_p(t)). \tag{A25}$$

The benchmark also optimizes event-probability gain over all distributions on the base support. The KL chain rule gives $\mathrm{KL}(P_0\|Q) \ge d_{\mathrm{B}}(p\|Q(A))$, with equality when Q preserves the conditional laws inside A and its complement. The event tilt has exactly these conditional laws, so its optimal gain equals the unrestricted event-probability optimum.

Part (a): local information and attainable gain. The score of the finite tilt family is $\partial_t \log P_t(s) = \mathbf{1}_A(s) - q_p(t)$. Its variance gives $F(t) = q_p(t)(1 - q_p(t)) = J'(t) > 0$, so the full natural gradient is one. The candidate $t_{\mathrm{B,cand}} = \log(1 + 1/(10p))$ has normalizer $11/10$, so $\mathcal{K}_p(t_{\mathrm{B,cand}}) = \log(11/10) - pt_{\mathrm{B,cand}} \le 1/10$ and $q_p(t_{\mathrm{B,cand}}) - p = (1 - p)/11 \ge 9/100 - p$. For the limit, the bound $\mathcal{K}_p(t) \ge (1 - p)t + \log p$ gives

$$0 \le pT(p) \le \frac{p/10 - p\log p}{1 - p} \longrightarrow 0. \tag{A26}$$

At the boundary, the normalizer equals $e^{1/10 + pT(p)}$. Hence

$$\Delta_p^\star = (1 - p)\big(1 - e^{-1/10 - pT(p)}\big) \longrightarrow 1 - e^{-1/10}. \tag{A27}$$

Replacing $1/10$ by any fixed $\eta > 0$ in this argument gives limiting gain $1 - e^{-\eta}$.

Part (b): separation and indistinguishable samples. We choose a step too short to be useful under the smaller mass and already infeasible under the larger mass, turning calibration into a two-law test. Choose $t_{\mathrm{sep}}(p) = \log(1/(16p))$ to separate the two laws: the factor 16 gives limiting event probability $1/17 < 27/400$ at mass p, but limiting KL $\log(9/8) > 1/10$ at mass $2p$. Usefulness requires $q_p(t) \ge 27/400 + p/4$, whereas for $p \le p_0/2$, $q_p(t_{\mathrm{sep}}(p)) = 1/(17 - 16p) \le 128/2175 < 3/50$. Thus every useful output at p exceeds $t_{\mathrm{sep}}(p)$. For the alternative mass, both terms in the expression below decrease on $(0, p_0]$, because their derivatives $-2/(9/8 - 2p)$ and $2(\log(16p) + 1)$ are negative there. Evaluating at p_0 therefore bounds the cost for every $p \le p_0/2$:

$$\mathcal{K}_{2p}(t_{\mathrm{sep}}(p)) = \log(9/8 - 2p) + 2p\log(16p) \ge \kappa_{\mathrm{sep}} := \log(575/512) - \frac{\log 64}{512} > 0.1079225. \tag{A28}$$

The fixed margin above the KL budget will also allow finite-precision thresholding in the computational reduction below. Every feasible output at $2p$ is therefore below $t_{\mathrm{sep}}(p)$. Bernoulli base laws on the same event $\{1\}$ have the same supplied vector, identically one. The laws $\mathrm{Bern}(p)$ and $\mathrm{Bern}(2p)$ have total variation distance p, and their useful-step sets are disjoint. Lemma A.1 therefore bounds the error sum below by $(1-p)^m$, while the two success guarantees bound it above by 2δ. Hence $(1-p)^m \le 2\delta$, proving the lower bound. At $\delta = 1/3$, $-\log(1-p) \le p/(1-p)$ gives $m \ge (1-p)\log(3/2)/p \ge 1/(3p)$ on the stated range.

Part (c): confidence calibration. We first show that the true mass belongs to the confidence interval with the required probability and that the confidence boundary defines a finite step for every sample outcome. The exponential Markov inequality, minimized over its parameter, gives $\mathbb{P}_p(\widehat{p} \ge x) \le e^{-m d_{\mathrm{B}}(x\|p)}$ for $x > p$. Negative parameters give the lower tail (Hoeffding, 1963). Monotonicity on each side and endpoint limits yield

$$\mathbb{P}\{m d_{\mathrm{B}}(\widehat{p}\|p) > \log(2/\delta)\} \le \delta. \tag{A29}$$

As a sublevel set of the lower-semicontinuous convex function $r \mapsto d_{\mathrm{B}}(\widehat{p}\|r)$, $\mathcal{I}_m$ is a compact interval. Because $\log(2/\delta) > 0$, it contains a mass in $(0,1)$ even at zero or full counts. The function $\mathcal{K}^{\max}_{\mathcal{I}_m}(t) := \max_{r\in\mathcal{I}_m} \mathcal{K}_r(t)$ is continuous, with $\mathcal{K}^{\max}_{\mathcal{I}_m}(0) = 0$, and diverges as $t \to \infty$. For $t > 0$, a maximizing mass lies in $(0,1)$, since $\mathcal{K}_0(t) = \mathcal{K}_1(t) = 0$ and $\mathcal{K}_r(t) > 0$ for $0 < r < 1$. Evaluating at such a maximizer shows that $\mathcal{K}^{\max}_{\mathcal{I}_m}$ is strictly increasing. It therefore has a unique finite $1/10$ boundary, which defines $\widehat{t}_m$ and can be bracketed by doubling and approximated by bisection. On the coverage event, this step is feasible for p. For implementation, $r \mapsto \mathcal{K}_r(t)$ is concave at each fixed $t > 0$. Its unconstrained maximum occurs at $r = 1/t - 1/(e^t - 1)$. Over $\mathcal{I}_m$, use this value if it lies in the interval, and otherwise the nearest endpoint. This evaluates the worst-case KL needed at each step of the scalar boundary search.

To establish usefulness, we bound the largest plausible mass and exhibit a step feasible for every mass up to that bound. Put $\Lambda_8 = \log(8/\delta) \ge \log(2/\delta)$, and condition on coverage. Applying $-\log x \ge 2(1-\sqrt{x})$ to the Bernoulli outcomes gives

$$d_{\mathrm{B}}(a\|b) \ge (\sqrt{a} - \sqrt{b})^2 + (\sqrt{1-a} - \sqrt{1-b})^2. \tag{A30}$$

Endpoint limits preserve this inequality, allowing infinite KL. For $r, p \in \mathcal{I}_m$, applying this bound to $(\widehat{p}, r)$ and $(\widehat{p}, p)$, then using the triangle inequality, gives $\sqrt{r} \le \sqrt{p} + 2\sqrt{\Lambda_8/m}$. We relax $2\sqrt{\Lambda_8/m}$ to $\sqrt{8\Lambda_8/m}$. The sample bound with coefficient 2048 ensures $\sqrt{8\Lambda_8/(mp)} \le 1/16$, yielding

$$p \le p_{\mathrm{U}} := \max \mathcal{I}_m \le (\sqrt{p} + \sqrt{8\Lambda_8/m})^2 \le (17/16)^2 p = 289p/256. \tag{A31}$$

The candidate $t_{\mathrm{B,conf}} = \log(1 + 1/(10p_{\mathrm{U}}))$ has $\mathcal{K}_r(t_{\mathrm{B,conf}}) \le \log(1 + r/(10p_{\mathrm{U}})) \le \log(11/10) < 1/10$ for every $r \le p_{\mathrm{U}}$. Hence $\widehat{t}_m \ge t_{\mathrm{B,conf}}$, and

$$\Delta_p(\widehat{t}_m) \ge \frac{p(1-p)}{10p_{\mathrm{U}} + p} \ge \frac{1 - p_0}{10(17/16)^2 + 1} > 0.08129. \tag{A32}$$

For binary entropy $H(p)$, feasible q satisfies $(1-p)\log(1/(1-q)) = \mathcal{K}_p + H(p) + p\log q \le 1/10 + H(p)$. Therefore

$$\Delta^\star_p \le 1 - \exp\left(-\frac{1/10 + H(p)}{1-p}\right) \le 1 - \exp\left(-\frac{1/10 + H(p_0)}{1-p_0}\right) < 0.103. \tag{A33}$$

The exponent magnitude increases with p, with derivative $(1/10 - \log p)/(1-p)^2$. For the final inequality, use $H(p_0) \le p_0 \log(1/p_0) + p_0$ and $\log 2 < 7/10$: the exponent magnitude is below 0.108, and $1 - e^{-0.108} \le 0.108 - 0.108^2/2 + 0.108^3/6 < 0.103$. Since $0.08129 > (3/4)0.103 = 0.07725$, both conditions for a useful step hold on the same coverage event. The guaranteed gain exceeds the required fractional gain by more than 0.004, so $\Delta'_p \le 1/4$ permits downward rational rounding by at most 10^{-3} while preserving both conditions. Rounding can be chosen nonnegative, and monotonicity of $\mathcal{K}_p$ preserves feasibility. □

Encoding and complexity conventions. We use the standard binary encoding of Boolean formulas. This gives a succinct model description because it specifies probabilities without listing the full state space. Input and output lengths are measured in bits. NP consists of decision problems whose yes

answers have polynomial-length certificates verifiable in polynomial time. BPP consists of decision problems solvable in randomized polynomial time with error at most one third on every input (Arora & Barak, 2009).

Proof of Corollary 3.4. We reduce SAT to calibration by constructing an event from the input formula and thresholding the step returned by a hypothetical calibration procedure. The threshold depends only on the formula's number of variables. We establish finite-precision separation and polynomial-length valid outputs.

We first specify the event used in the main-text corollary. Draw $B_{\mathrm{tag}} \in \{0,1\}^{11}$ and $X \in \{0,1\}^n$ independently and uniformly, and define

$$A_\varphi = \{B_{\mathrm{tag}} = 0^{11}, \varphi(X) = 1\} \cup \{B_{\mathrm{tag}} = 1^{11}, X = 0^n\}. \tag{A34}$$

The tag bits separate the satisfying assignments from one additional accepted state. Their number ensures that even $2^n + 1$ accepted states have total mass below $p_0 = 2^{-10}$, as required by the calibration bounds. The two blocks are disjoint, so this event has mass $p = K/D_{\mathrm{B}}$, where $K = 1 + \#\mathrm{SAT}(\varphi)$, $D_{\mathrm{B}} = 2^{n+11}$, and $p < p_0$. This identity describes the mass. We construct the event and its base sampler without computing K. Set $t_{\mathrm{sep,B}} = (n+7)\log 2 = \log(D_{\mathrm{B}}/16)$. When $K = 1$, useful outputs have event probability at least $27/400 + p/4$, while $q_p(t_{\mathrm{sep,B}}) < 3/50$. Since $q_p' \le 1/4$, the probability gap $27/400 - 3/50 = 3/400$ gives $t > t_{\mathrm{sep,B}} + 4(3/400) = t_{\mathrm{sep,B}} + 0.03$.

For $K \ge 2$, $p \in [2/D_{\mathrm{B}}, p_0]$. Unlike the sampling pair, the satisfiable case allows many masses, so separation must hold throughout $[2/D_{\mathrm{B}}, p_0]$. Concavity in the mass, from

$$\partial_r \mathcal{K}_r(t) = \frac{e^t - 1}{1 + r(e^t - 1)} - t, \qquad \partial_r^2 \mathcal{K}_r(t) = -\frac{(e^t - 1)^2}{(1 + r(e^t - 1))^2} < 0, \tag{A35}$$

reduces the minimum at $t = t_{\mathrm{sep,B}}$ to the endpoints. At $2/D_{\mathrm{B}}$, (A28) gives cost at least κ_{sep}. At p_0, since $D_{\mathrm{B}} \ge 2^{14}$,

$$\mathcal{K}_{p_0}(t_{\mathrm{sep,B}}) \ge \mathcal{K}_{p_0}(\log 1024) = \log(2047/1024) - \log(1024)/1024 > 0.685 > \kappa_{\mathrm{sep}}. \tag{A36}$$

Thus any feasible output satisfies $t_{\mathrm{sep,B}} - t \ge \mathcal{K}_p(t_{\mathrm{sep,B}}) - \mathcal{K}_p(t) \ge \kappa_{\mathrm{sep}} - 1/10 > 0.0079$, using $0 \le \mathcal{K}_p' \le 1$. A polynomial-time rational approximation to $t_{\mathrm{sep,B}}$ within 10^{-3} separates these cases. For example, the convergent expansion $\log 2 = 2\sum_{j\ge 0} 3^{-(2j+1)}/(2j+1)$ needs only $\mathcal{O}(\log(n+7))$ terms to achieve error below $10^{-3}/(n+7)$, using rational arithmetic of polynomial bit length.

Useful steps of polynomial bit length exist. The step $t = \log(1 + 1/(10p)) = \mathcal{O}(n)$ has cost at most $\log(11/10) < 1/10$ and gain $(1-p)/11 \ge (1-p_0)/11 > 0.09082 > 0.07725$. Downward rounding within 10^{-3} preserves feasibility and usefulness. The rounded step needs only $\mathcal{O}(\log(n+1))$ integer bits and a constant number of fractional bits. This existence argument uses p, which is not supplied to the procedure. The predicate is polynomial-time constructible and uses at most one formula evaluation, and its base sampler uses $n + 11$ fair bits. Fresh fair bits simulate every draw the procedure takes, and the supplied natural gradient is one for every formula. Run a hypothetical calibration procedure once on this input. Comparing its returned step with the rational threshold decides the original formula's satisfiability, giving the stated randomized implication (Arora & Barak, 2009). □

For the confidence rule in Theorem B.2(c), at $\delta = 0.05$ the sufficient bound is $m \ge 10394/p$, about 1.04×10^9 draws at $p = 10^{-5}$.

Dependence on divergence orientation. The positive limiting gain above depends on the orientation of KL. The following bounds isolate that dependence by examining feasible gain before asking how many observations calibration would require. For an event tilt, let $q = P_t(A) \ge p = P_0(A)$. Because the conditional laws within A and its complement are unchanged, update-to-base KL is $d_{\mathrm{B}}(q\|p)$. The binary entropy bound gives

$$d_{\mathrm{B}}(q\|p) \ge q\log(1/p) - \log 2. \tag{A37}$$

At any fixed budget η, feasibility therefore requires $q \le (\eta + \log 2)/\log(1/p)$, so the optimal gain tends to zero as $p \downarrow 0$. With $\chi^2(Q\|P) = \mathbb{E}_P[(dQ/dP - 1)^2]$, direct summation gives

$$\chi^2(P_t\|P_0) = \frac{(q-p)^2}{p(1-p)}, \qquad q - p \le \sqrt{\eta p(1-p)} \quad \text{when } \chi^2(P_t\|P_0) \le \eta. \tag{A38}$$

Thus the nonvanishing-gain conclusion in Theorem B.2(a) does not extend to either of these constraints. These comparisons do not determine their optimal calibration sample rates.

B.3 Dependence on the required improvement

We now quantify how the sampling cost grows as the allowed gain loss shrinks. The explicit bounds below imply the order statement in Theorem 3.3. Use the event-tilt model and sampling access defined in Appendix B.2. At event mass r, a step $t \geq 0$ is useful if $\mathcal{K}_r(t) \leq 1/10$ and $\Delta_r(t) \geq (1-\varepsilon)\Delta_r^\star$. Fix $p_0 = 2^{-10}$, $0 < p \leq p_0/2$, $0 < \varepsilon \leq 1/4$ and $0 < \delta \leq 1/3$. Any procedure that uses at most m draws and returns a useful step with probability at least $1-\delta$ under both Bernoulli masses p and $(1+4\varepsilon)p$ requires

$$m \geq \frac{(1-p_0)(1-2\delta)\log((1-\delta)/\delta)}{16p\varepsilon^2}. \tag{A39}$$

Conversely, apply the confidence rule in Theorem B.2(c). Round its output downward by at most $\varepsilon/32$ to a nonnegative step. This step is useful at mass p with probability at least $1-\delta$ whenever

$$m \geq \frac{1024\log(2/\delta)}{p\varepsilon^2}. \tag{A40}$$

At fixed confidence, the matching worst-case rate corresponds to an expected event count $mp = \Theta(\varepsilon^{-2})$, even with the exact natural gradient supplied.

Proof of Theorem 3.3. For the lower bound, a relative change in mass moves the KL boundary farther than the width of the useful-step interval. For the upper bound, the same sensitivity estimate converts mass-estimation error into gain loss, with a separate allowance for rounding.

Width and motion of the useful-step interval. Retain $p_0 = 2^{-10}$, $0 < p \leq p_0/2$, $0 < \varepsilon \leq 1/4$, and $0 < \delta \leq 1/3$ from the theorem. Let $q_p^\star > p$ solve $d_{\mathrm{B}}(p\|q_p^\star) = 1/10$, and put $\alpha = 1-\varepsilon$. Inverting the event-probability map gives the largest feasible step and the first step attaining the required gain:

$$T(p) = \mathrm{logit}(q_p^\star) - \mathrm{logit}(p), \qquad L_\varepsilon(p) = \mathrm{logit}(p + \alpha\Delta_p^\star) - \mathrm{logit}(p). \tag{A41}$$

A step is useful exactly when it lies in $[L_\varepsilon(p), T(p)]$. The proof of Theorem B.2 gives $0.09 - p \leq \Delta_p^\star < 0.103$ and $q_p^\star < 0.104$ for $p \leq p_0$. At $t = \log(1/p)$, $\mathcal{K}_p(t) = \log(2-p) - p\log(1/p) > \log(3/2) - 7/1024 > 1/10$, so $T(p) < \log(1/p)$. Since $\mathcal{K}_p'(T(p)) = \Delta_p^\star > 0$, implicit differentiation of $\mathcal{K}_p(T(p)) = 1/10$ yields

$$-\frac{\mathrm{d}T}{\mathrm{d}\log p} = \frac{1}{1-p} - \frac{pT(p)}{\Delta_p^\star}, \qquad 0.923 < -\frac{\mathrm{d}T}{\mathrm{d}\log p} \leq \frac{1}{1-p_0}. \tag{A42}$$

For the lower inequality, $p\log(1/p)$ increases up to p_0 and is smaller than $7/1024$ there, while $(7/1024)/(0.09 - p_0) = 700/9116 < 0.077$. Integrating the derivative of logit between the two event-probability targets gives

$$T(p) - L_\varepsilon(p) = \int_\alpha^1 \frac{\Delta_p^\star \, \mathrm{d}u}{(p + u\Delta_p^\star)(1 - p - u\Delta_p^\star)} \leq \frac{\varepsilon}{\alpha(1-q_p^\star)} < 1.489\varepsilon. \tag{A43}$$

Choose $\chi_\varepsilon = 1+4\varepsilon \leq 2$ so that changing the mass by this factor moves the boundary by more than the width of the useful-step interval. Since $\log(1+4\varepsilon) \geq 2\varepsilon$, Equation (A42) implies $T(p) - T(\chi_\varepsilon p) > 1.846\varepsilon$. Consequently $L_\varepsilon(p) > T(\chi_\varepsilon p)$: every useful output at p exceeds every feasible output at $\chi_\varepsilon p$.

Sample complexity lower bound. Thresholding the output between these endpoints selects p with probability at least $1-\delta$ under p and at most δ under $\chi_\varepsilon p$. By Lemma A.1, the law of this test outcome has KL divergence at most m times the single-draw value. Monotonicity in the two test probabilities then gives

$$m\, d_{\mathrm{B}}(p\|\chi_\varepsilon p) \geq (1-2\delta)\log\frac{1-\delta}{\delta}. \tag{A44}$$

Using $\log x \leq x - 1$ in both terms of binary KL gives $d_{\mathrm{B}}(a\|b) \leq (a-b)^2/[b(1-b)]$ for $0 < a, b < 1$. Hence $d_{\mathrm{B}}(p\|\chi_\varepsilon p) \leq (\chi_\varepsilon p - p)^2/[\chi_\varepsilon p(1-\chi_\varepsilon p)] \leq 16p\varepsilon^2/(1-p_0)$ proves Equation (A39). The supplied natural gradient 1 is identical under both laws.

Confidence calibration and gain loss. For the upper bound, fix a sample outcome with $p \in \mathcal{I}_m$. Set $\Lambda_2 = \log(2/\delta)$ and let $p_{\mathrm{U}} = \max \mathcal{I}_m$ be the upper endpoint of the nonempty compact confidence interval, so $p_{\mathrm{U}} \geq p > 0$. The coefficient $1024 = (2 \cdot 16)^2$ makes the relative square-root error at most $\varepsilon/16$:

$$\sqrt{p_{\mathrm{U}}} \leq \sqrt{p} + 2\sqrt{\Lambda_2/m}, \qquad m \geq \frac{1024\Lambda_2}{p\varepsilon^2} \implies p_{\mathrm{U}} \leq (1+\varepsilon/16)^2 p \leq (1+\varepsilon/4)p < p_0. \tag{A45}$$

For fixed positive t, $\partial \mathcal{K}_r(t)/\partial r$ decreases in r. At $r = p_{\mathrm{U}}$ and $t = T(p_{\mathrm{U}})$ it equals $-\Delta^\star_{p_{\mathrm{U}}} T'(p_{\mathrm{U}}) > 0$ by Equation (A42). Hence $\mathcal{K}_r(T(p_{\mathrm{U}})) \leq \mathcal{K}_{p_{\mathrm{U}}}(T(p_{\mathrm{U}})) = 1/10$ for every $0 \leq r \leq p_{\mathrm{U}}$. The confidence rule therefore chooses $\widehat{t}_m \geq T(p_{\mathrm{U}})$ and, on the coverage event, $\widehat{t}_m \leq T(p)$. Applying Equation (A42) once more gives

$$T(p) - T(p_{\mathrm{U}}) \leq \frac{\log(p_{\mathrm{U}}/p)}{1-p_0} \leq \frac{\varepsilon}{4(1-p_0)}. \tag{A46}$$

Since $0 \leq \Delta'_p(t) \leq 1/4$ and $\Delta^\star_p \geq 0.09 - p_0$, the selected step and downward rounding by at most $\varepsilon/32$ together incur relative gain loss at most

$$\varepsilon \left[\frac{1}{16(1-p_0)(0.09-p_0)} + \frac{1}{128(0.09-p_0)} \right] < 0.8\varepsilon < \varepsilon. \tag{A47}$$

The first bracketed term accounts for conservative calibration and the second for rounding. Round downward to a nonnegative step. This rounding preserves feasibility because $\mathcal{K}_p$ increases on $[0, \infty)$. Coverage has probability at least $1 - \delta$, proving Equation (A40). This rounding budget must scale with ε to preserve the stated accuracy. For $0 < \delta \leq 1/3$, $1 - 2\delta \geq 1/3$ and $(1-\delta)^2 \geq \delta$, so

$$\frac{1}{6} \log \frac{1}{\delta} \leq (1-2\delta) \log \frac{1-\delta}{\delta} \leq \log \frac{1}{\delta}. \tag{A48}$$

Thus the testing lower bound has the same logarithmic confidence dependence as the constructive upper bound, proving the uniform order statement. □

C Quadratic rescaling and finite KL

We separate Fisher matrix estimation error from error in the quadratic approximation to finite-step KL. The scalar example with exact Fisher information isolates approximation error, which is distinct from sampling error in the learned-head KL estimates. Let $F, \widehat{F} \succeq 0$ be the population and estimated Fisher information matrices.

For $v^\top \widehat{F} v > 0$ and $\eta > 0$, the proposed step and its population quadratic cost satisfy

$$\Delta\theta_\eta = \sqrt{\frac{2\eta}{v^\top \widehat{F} v}}\, v, \qquad \frac{\Delta\theta_\eta^\top F \Delta\theta_\eta}{2\eta} = \frac{v^\top F v}{v^\top \widehat{F} v}. \tag{A49}$$

For a chosen budget fraction $0 < \chi_{\mathrm{bud}} \leq 1$, targeting $\chi_{\mathrm{bud}}\eta$ under budget η multiplies this ratio by χ_{bud}. Scaling both budgets by χ_{bud} leaves it unchanged. This identity describes error within the quadratic approximation.

Even with exact scalar Fisher information, a quadratic proposal can violate the finite-step KL budget. For the binary tilt at $p = 1/2048$ and budget 0.1, the quadratic step $t = \sqrt{0.2/[p(1-p)]} \approx 20.24$ has true KL 12.609. These values follow by substituting the step into

$$\mathcal{K}_p(t) = \log(1 - p + pe^t) - pt. \tag{A50}$$

A 10% quadratic-budget margin scales the step by $\sqrt{0.9}$ and still gives KL 11.571. Thus reserving budget in the quadratic approximation does not ensure finite-KL feasibility in this example. Section 3.1 gives a four-state example in which all laws have the same quadratic step. That step is about 470 times longer than their KL-feasible boundary steps.

D DIRECTION RECOVERY IN LOGISTIC MODELS

We first prove Theorem 3.5 by showing that a coordinate of the normalized direction separates unsatisfiable from satisfiable formulas. We then prove Corollaries D.1 and D.2 by separating the first coordinates of useful finite updates. The second argument establishes hardness directly from the returned update. Appendix A describes the anchor state, tag bits and branch bits used below.

Encoding conventions. For a formula φ with code $c = \langle\varphi\rangle$, the proofs index the construction by c: $Q_c = Q_\varphi$, $F_c^{\mathrm{L}} = F_\varphi^{\mathrm{L}}$, $v_c^{\mathrm{L}} = v_\varphi^{\mathrm{L}}$, $\nu_c^{\mathrm{L}} = \nu_\varphi^{\mathrm{L}}$ and $M(c) = \#\mathrm{SAT}(\varphi)$. The count describes the resulting law. We construct its sampler by evaluating assignments, without computing $M(c)$. Malformed formula codes have $M(c) = 0$, and decision languages reject malformed outer encodings and $n < 3$.

Proof of Theorem 3.5. **Sampler construction and local geometry.** The sampler must implement the covariate law by checking assignments, without counting them. The anchor state ensures a positive definite Fisher matrix even for an unsatisfiable formula. Write $\mathsf{E}(c, x)$ for evaluation of the formula encoded by c on $x \in \{0,1\}^n$, with value zero for malformed formula codes. Three auxiliary tag bits distinguish the formula-evaluation block, the anchor state and a separate fixed state. Two additional branch bits select which covariate value to produce. Together with the assignment bits, these bits give $D_{\mathrm{L}} = 2^{n+5}$ equally likely inputs. Let $W = (a, x)$ be uniform on $\Omega_n^{\mathrm{L}} = \{0,1\}^3 \times \{0,1\}^n$. Let $a_c(a, x)$ be the indicator of $(a = 000, \mathsf{E}(c, x) = 1)$ or $(a = 111, x = 0^n)$. These disjoint cases contain $K = M(c) + 1$ states. With $w_0 = (001, 0^n)$ and an independent uniform $B_{\mathrm{br}} \in \{0,1,2,3\}$, set

$$X_c(B_{\mathrm{br}}, W) = \begin{cases} e_1, & B_{\mathrm{br}} = 0,\ a_c(W) = 1, \\ e_2, & B_{\mathrm{br}} = 1,\ a_c(W) = 1, \\ (1,1)^{\mathsf{T}}, & B_{\mathrm{br}} = 2,\ W = w_0, \\ 0, & \text{otherwise.} \end{cases} \tag{A51}$$

The branch counts are $K, K, 1, D_{\mathrm{L}} - 2K - 1$. The last is positive because $K \le 2^n + 1$ and $2^{n+1} + 3 < 2^{n+5} = D_{\mathrm{L}}$. The resulting law is Q_c, sampled using $n + 5$ fair bits and at most one formula evaluation. At $\theta = 0$, one extra independent fair bit samples the label.

The score is $(Y - \mathrm{sigm}(\theta^{\mathsf{T}} X))X$. Averaging over model labels gives the Fisher matrix $\mathbb{E}_Q[\mathrm{sigm}(\theta^{\mathsf{T}} X)(1 - \mathrm{sigm}(\theta^{\mathsf{T}} X))XX^{\mathsf{T}}]$. At zero this is $\mathbb{E}_Q[XX^{\mathsf{T}}]/4$, the matrix in (5), with eigenvalues $K/(4D_{\mathrm{L}})$ and $(K+2)/(4D_{\mathrm{L}})$. The query gradient $e_1/2$ gives $v_c^{\mathrm{L}} = \frac{2D_{\mathrm{L}}}{K(K+2)}(K+1, -1)^{\mathsf{T}}$, whose normalization proves (a). The query-only Fisher matrix $e_1 e_1^{\mathsf{T}}/4$ is singular.

Computational reduction. Given a formula, construct the sampler above and run a hypothetical direction-recovery procedure on $(1^n, c)$. We now show how to decide satisfiability from its returned direction. The second coordinate of the unit direction is $-1/\sqrt{5}$ for $K = 1$ and at least $-1/\sqrt{10}$ for $K \ge 2$. Thus $(\nu_c^{\mathrm{L}})_2 > -3/8$ exactly when c is satisfiable. A verifier checks well-formedness and a satisfying assignment. Renumbering variables and padding to $n \ge 3$ preserves satisfiability, giving a polynomial many-one reduction and proving (d). Since $-1/\sqrt{5} + 1/32 < -3/8 < -1/\sqrt{10} - 1/32$, every allowed approximate output also decides SAT by a rational comparison, proving (b) (Arora & Barak, 2009). The description has polynomial size, v_c^{L} has $\mathcal{O}(n)$-bit rational coordinates, and constant-accuracy rational unit-vector approximations exist without requiring exact square roots. The sampler simulates adaptive independent draws with fresh bits, so a polynomial-time recovery procedure would imply NP $\subseteq$ BPP.

Sample complexity lower bound. For (c), withhold codes of a contradiction and a conjunction fixing all variables. Their $K = 1, 2$ laws in (4) have total variation distance $2/D_{\mathrm{L}}$ per observation, unchanged by the independent fair label. Thresholding the estimated second coordinate at $-3/8$ gives

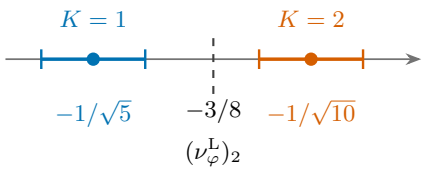

Figure A1: Exact second coordinates of normalized directions with error intervals of radius $1/32$. The threshold $-3/8$ separates unsatisfiable ($K = 1$) from satisfiable ($K \ge 2$) formulas. The coordinate increases with K.

a test with error at most $1/3$ under each law. By Lemma A.1, its error sum is at least $1 - 2m/D_{\mathrm{L}}$. Hence $m \geq D_{\mathrm{L}}/6$, even with unrestricted computation and known n. □

Corollary D.1 (Finite KL-constrained query improvement)**.** *Fix an integer $n \geq 3$ and a formula code c on n variables. Write $M(c)$ for its number of satisfying assignments, set $K = M(c) + 1$ and $D_{\mathrm{L}} = 2^{n+5}$, and let e_1, e_2 be the standard basis of $\mathbb{R}^2$. The covariate law satisfies*

$$Q_c(e_1) = Q_c(e_2) = K/D_{\mathrm{L}}, \qquad Q_c(e_1 + e_2) = 1/D_{\mathrm{L}}, \qquad Q_c(0) = 1 - (2K+1)/D_{\mathrm{L}}. \tag{A52}$$

For $\theta \in \mathbb{R}^2$, let $\pi_\theta(Y = 1 \mid X) = \mathrm{sigm}(\theta^{\mathsf{T}} X)$, where $\mathrm{sigm}(s) = (1 + e^{-s})^{-1}$, and define

$$\mathcal{K}_c^{\mathrm{L}}(\theta) = \mathbb{E}_{Q_c} \mathrm{KL}(\pi_0(\cdot \mid X) \| \pi_\theta(\cdot \mid X)), \quad \eta_{\mathrm{L}} = \frac{1}{100 D_{\mathrm{L}}},$$
$$\ell_c^* = \max_{\theta \in \mathbb{R}^2 : \mathcal{K}_c^{\mathrm{L}}(\theta) \leq \eta_{\mathrm{L}}} \log \mathrm{sigm}(\theta_1). \tag{A53}$$

The input is $(1^n, c)$, with n encoded in unary. Neither K, the population Fisher matrix nor the natural gradient is supplied. Returning a polynomial-length rational $\widehat{\theta} \in \mathbb{Q}^2$ with $\mathcal{K}_c^{\mathrm{L}}(\widehat{\theta}) \leq \eta_{\mathrm{L}}$ and $\log \mathrm{sigm}(\widehat{\theta}_1) \geq \ell_c^* - 1/200$ is NP-hard under polynomial-time Turing reductions. Unless $\mathrm{NP} \subseteq \mathrm{BPP}$, no randomized polynomial-time procedure succeeds on every input with probability at least $2/3$, even with adaptive independent exact draws (X, Y) from $Q_c \pi_0$.

Proof of Corollary D.1. We compare the first coordinates attainable in the two cases, then translate their separation into an objective gap. Multiplying the cost by the common denominator removes the rare-covariate scale.

Exact KL cost. Let $\phi(t) = \log \cosh(t/2)$. The identity $\mathrm{KL}(\mathrm{Bern}(1/2) \| \mathrm{Bern}(\mathrm{sigm}(t))) = \phi(t)$ gives the exact cost

$$D_{\mathrm{L}} \mathcal{K}_c^{\mathrm{L}}(\theta) = K\phi(\theta_1) + K\phi(\theta_2) + \phi(\theta_1 + \theta_2). \tag{A54}$$

Since $K \geq 1$ and $\phi(t) \geq |t|/2 - \log 2$, the cost tends to infinity as $\|\theta\|_2 \to \infty$. Its nonempty sublevel set at η_{L} is therefore compact, so the continuous objective attains its maximum. Since $\phi(0) = \phi'(0) = 0$ and $\phi'' = (1 - \tanh^2(t/2))/4$, integration gives

$$\phi(t) \leq t^2/8 \quad (t \in \mathbb{R}), \qquad \phi(t) \geq 3t^2/32 \quad (|t| \leq 1). \tag{A55}$$

The lower bound uses $|\tanh(t/2)| \leq |t|/2$.

Separation of useful outputs. We exhibit one feasible point for the unsatisfiable case and bound every feasible first coordinate in the satisfiable case. At $K = 1$, $(t, -t/2)$ with $t = 23/100$ has scaled cost at most $3t^2/16 < 1/100$.

For $K \geq 2$, a feasible positive $t = \theta_1$ obeys $2\phi(t) \leq 1/100$, hence $t < 1$. For fixed $t > 0$, the function $h_t(r) = K\phi(r) + \phi(t + r)$ is strictly convex and coercive, with $h_t'(-t) = K\phi'(-t) < 0 < h_t'(0) = \phi'(t)$. Its unique minimizer r therefore lies in $(-t, 0)$. Replacing θ_2 by r can only decrease the cost, and all three arguments $t, r, t + r$ then lie in $[-1, 1]$. Thus

$$D_{\mathrm{L}} \mathcal{K}_c^{\mathrm{L}}(\theta) \geq \frac{3}{32}\{2t^2 + 2r^2 + (t + r)^2\} = \frac{9}{32}(r + t/3)^2 + \frac{t^2}{4}. \tag{A56}$$

Thus every feasible positive first coordinate is at most $1/5$. Nonpositive coordinates satisfy the same upper bound. On $[0, 23/100]$, $e^t \leq (1 - t)^{-1} \leq 100/77 < 3/2$, so $(\log \mathrm{sigm})'(t) > 2/5$ and

$$\log \mathrm{sigm}(23/100) - \log \mathrm{sigm}(43/200) > 3/500 > 1/200. \tag{A57}$$

A valid output has $\widehat{\theta}_1 > 43/200$ for $K = 1$ and $\widehat{\theta}_1 \leq 1/5$ for $K \geq 2$. Thus a hypothetical finite-update procedure would decide satisfiability by thresholding its returned first coordinate. The reduction need not evaluate the output's KL cost or objective value.

Polynomial-length rational outputs. To obtain polynomial-length rational outputs, first observe that every feasible coordinate lies in $(-1, 1)$. Each contributes at least $\phi(\theta_i)$ to the scaled cost, while $\phi(1) \geq 3/32 > 1/100$ and ϕ increases with absolute value. Choose a shrinkage $\chi_{\mathrm{shrink}} = 10^{-3}$ and scale an optimizer by $1 - \chi_{\mathrm{shrink}}$. Convexity and $\mathcal{K}_c^{\mathrm{L}}(0) = 0$ leave scaled-KL slack at least $\chi_{\mathrm{shrink}}/100 = 10^{-5}$. Since $|(\log \mathrm{sigm})'| \leq 1$, the objective loss is below χ_{shrink}. Round each coordinate to a rational with a power-of-two denominator within $1/(10^6 D_{\mathrm{L}})$, requiring $\mathcal{O}(n)$

fractional bits. Since $|\phi'| \le 1/2$, scaled-cost error is at most $(K+1)/(10^6 D_{\mathrm{L}}) \le 10^{-6}$. The rounding cost is below the available slack, and total objective loss is below $10^{-3}+10^{-6} < 1/200$. The separating threshold and exact-draw simulation therefore give both hardness claims with polynomial-length outputs. □

Corollary D.2 (Sample-only finite KL-constrained improvement)**.** *Fix an integer $n \ge 3$, set $D_{\mathrm{L}} = 2^{n+5}$, and use the two covariate laws with $K = 1, 2$ in Corollary D.1. Use the same logistic model, KL cost $\mathcal{K}_c^{\mathrm{L}}$, and optimum ℓ_c^* defined there. The learner knows n and the common budget $\eta_{\mathrm{L}} = 1/(100 D_{\mathrm{L}})$, but not the codes, the population Fisher matrix or the natural-gradient direction. Consider any procedure using at most m independent draws (X, Y) at $\theta = 0$ and returning $\widehat{\theta} \in \mathbb{R}^2$. Suppose its output satisfies*

$$\mathcal{K}_c^{\mathrm{L}}(\widehat{\theta}) \le \eta_{\mathrm{L}}, \qquad \log \mathrm{sigm}(\widehat{\theta}_1) \ge \ell_c^* - 1/200 \tag{A58}$$

with probability at least $2/3$ under each law. Then $m \ge D_{\mathrm{L}}/6$, regardless of computational power. The same bound holds if the objective requirement is replaced by

$$\log \mathrm{sigm}(\widehat{\theta}_1) + \log 2 \ge \frac{99}{100}(\ell_c^* + \log 2). \tag{A59}$$

Proof of Corollary D.2. Reuse the preceding output separation to test the two observation laws. The proof of Corollary D.1 gives $\widehat{\theta}_1 > 43/200$ on success at $K = 1$, whereas feasibility at $K = 2$ requires $\widehat{\theta}_1 \le 1/5$. Thresholding at $43/200$ therefore tests the two laws with error at most $1/3$ under each. The total variation distance for one observation is $2/D_{\mathrm{L}}$. By Lemma A.1, the testing error sum is at least $1 - 2m/D_{\mathrm{L}}$, yielding $m \ge D_{\mathrm{L}}/6$.

For the relative-gain requirement, feasibility implies $|\theta_1| < 1$, because $\phi(1) \ge 3/32 > 1/100$. A sufficiently small positive first-coordinate step is feasible, and $(\log \mathrm{sigm})'(t) \le 1/2$ for $t \ge 0$, so $0 < \ell_c^* + \log 2 < 1/2$. Attaining $99/100$ of this gain leaves objective error below $1/200$, to which the first claim applies. □

E Damped natural-gradient estimation and lower bounds

Proposition E.1 and Corollary E.3 establish the upper and lower bounds summarized in Proposition 3.6. We then quantify the bias introduced by damping and identify when it makes the rare-event contribution negligible at the requested accuracy.

E.1 Fisher matrix estimation

Differences of independent observations let us estimate covariance without knowing the population mean. We measure matrix error relative to the damped population matrix and apply a concentration inequality. The resulting spectral bounds control the error in the estimated inverse.

Proposition E.1 (Estimating a damped natural gradient from samples)**.** *Let P be a probability law on $\mathcal{X}$ and let $h : \mathcal{X} \to [0,1]^d$ be a known measurable statistic. Set $F = \mathrm{Cov}_P(h)$. Fix a known $\lambda > 0$, parameters $0 < \rho, \delta < 1$, and integers $d, s \ge 1$. From $2s$ independent draws from P, form $\xi_i = (h(S_{2i}) - h(S_{2i-1}))/\sqrt{2}$, $\widehat{F} = s^{-1}\sum_i \xi_i \xi_i^{\mathsf{T}}$, $\mathsf{A} = F + \lambda I_d$, and $\widehat{\mathsf{A}} = \widehat{F} + \lambda I_d$. If*

$$s \ge \frac{3}{\rho^2}\left(1 + \frac{d}{2\lambda}\right)\log\frac{2d}{\delta}, \tag{A60}$$

then with probability at least $1 - \delta$,

$$\begin{aligned} (1-\rho)\mathsf{A} \preceq \widehat{\mathsf{A}} \preceq (1+\rho)\mathsf{A}, \\ \|\widehat{\mathsf{A}}^{-1} b - \mathsf{A}^{-1} b\|_{\mathsf{A}} \le \frac{\rho}{1-\rho}\|\mathsf{A}^{-1} b\|_{\mathsf{A}} \quad \text{for all } b \in \mathbb{R}^d, \end{aligned} \tag{A61}$$

where $\|z\|_{\mathsf{A}}^2 = z^{\mathsf{T}}\mathsf{A}z$. For the uncentered moment $F = \mathbb{E}_P hh^{\mathsf{T}}$, the same guarantees hold for $\widehat{F} = m^{-1}\sum_{i=1}^m h(S_i)h(S_i)^{\mathsf{T}}$ from m independent draws whenever $m \ge 3\rho^{-2}(1 + d/\lambda)\log(2d/\delta)$.

This specializes bounds for regularized covariance estimation and inversion (Hsu et al., 2014). Taking $\rho = \varepsilon/(1+\varepsilon)$ and counting $m = 2s$ individual draws gives the upper bound in Proposition 3.6.

Proof of Proposition E.1. **Normalized matrix concentration.** Independence gives $\mathbb{E}\xi_i\xi_i^\mathsf{T} = F$, and bounded coordinates give $\|\xi_i\|_2^2 \le d/2$. The independent matrices

$$\mathsf{W}_i = \mathsf{A}^{-1/2}(\xi_i\xi_i^\mathsf{T} + \lambda I_d)\mathsf{A}^{-1/2} \tag{A62}$$

have mean I_d and largest eigenvalue at most $R_\lambda = 1 + d/(2\lambda)$. Matrix Chernoff (Tropp, 2012) gives

$$\mathbb{P}\{\|s^{-1}\sum_i \mathsf{W}_i - I_d\|_2 > \rho\} \le 2de^{-s\rho^2/(3R_\lambda)} \le \delta. \tag{A63}$$

Uniform inverse error. On the complement of the event in Equation (A63), $\mathsf{B} = \mathsf{A}^{-1/2}\widehat{\mathsf{A}}\mathsf{A}^{-1/2}$ has spectrum in $[1-\rho, 1+\rho]$ and $\|\mathsf{B}^{-1} - I_d\|_2 \le \rho/(1-\rho)$. Using

$$\mathsf{A}^{1/2}(\widehat{\mathsf{A}}^{-1}b - \mathsf{A}^{-1}b) = (\mathsf{B}^{-1} - I_d)\mathsf{A}^{-1/2}b \tag{A64}$$

proves both bounds simultaneously, even when b is chosen from the sample. The matrix normalization is used only in the proof. The estimator does not require the population matrix. Both linear systems use the same right-hand side, so the bound isolates error from estimating the Fisher matrix. For uncentered moments, $\|h\|^2 \le d$ gives the replacement constant. The base-point logistic Fisher information matrix is obtained with $h = X/2$. □

Proposition E.2 (Scalar lower bound under damping)**.** *Fix a known $0 < \lambda \le 1/8$ and $0 < \rho \le 1/16$. Let $G_\lambda(p) = [p(1-p) + \lambda]^{-1}$. Consider any sample-only procedure using at most m independent Bernoulli draws and returning $\widehat{G}$. Suppose its output satisfies $|\widehat{G} - G_\lambda(p)| \le \rho G_\lambda(p)$ with probability at least $2/3$ for every Bernoulli law with $p \in (0,1)$. Then $m \ge 1/(360\lambda\rho^2)$ observations are necessary in the worst case.*

Proof of Proposition E.2. Choose event probabilities on the damping scale to separate accurate inverse estimates while keeping the observation laws close. To separate the relative-error intervals, take $p_- = \lambda, p_+ = (1+8\rho)\lambda$ and $I_\pm = p_\pm(1-p_\pm)$. Then $p_+ \le 3/16$, $I_- + \lambda \le 2\lambda$, and $I_+ - I_- = 8\rho\lambda(1 - p_- - p_+) \ge (11/2)\rho\lambda$. Consequently

$$\frac{G_\lambda(p_-)}{G_\lambda(p_+)} \ge 1 + \frac{11}{4}\rho > \frac{1+\rho}{1-\rho}. \tag{A65}$$

The accuracy intervals are disjoint, so a sufficiently accurate estimate identifies which law generated the observations. The bound $\log x \le x - 1$ gives

$$d_{\mathrm{B}}(p_-\|p_+) \le \frac{(p_+ - p_-)^2}{p_+(1-p_+)} \le 80\lambda\rho^2. \tag{A66}$$

By Lemma A.1 and Pinsker's inequality (van Erven & Harremoës, 2014), the output laws under the two masses are within total variation $\sqrt{40m\lambda\rho^2}$. Testing with both errors at most $1/3$ requires total variation at least $1/3$. Thus $40m\lambda\rho^2 \ge 1/9$, giving $360 = 9 \cdot 40$ in the denominator. □

Corollary E.3 (Dimension dependence under coordinate-wise boundedness)**.** *For an integer $d \ge 1$, a fixed known $0 < \lambda \le d/8$, and $0 < \varepsilon \le 1/16$, there exist a known statistic $h : \{0,1\} \to [0,1]^d$ and a fixed known unit vector b with the following property. Write $F = \mathrm{Cov}_P(h)$ and $\mathsf{A} = F + \lambda I_d$. Consider a sample-only procedure returning $\widehat{v} \in \mathbb{R}^d$ from at most m independent draws. If its output satisfies*

$$\mathbb{P}_P\{\|\widehat{v} - \mathsf{A}^{-1}b\|_\mathsf{A} \le \varepsilon\|\mathsf{A}^{-1}b\|_\mathsf{A}\} \ge 2/3, \tag{A67}$$

for every Bernoulli law P, then $m \ge d/(360\lambda\varepsilon^2)$.

Proof of Corollary E.3. Embed the scalar problem along one known direction and project any accurate vector estimate onto that direction. For $Z_{\mathrm{Bern}} \sim \mathrm{Bern}(p)$, set $h(Z_{\mathrm{Bern}}) = Z_{\mathrm{Bern}}\mathbf{1}_d$, $b = \mathbf{1}_d/\sqrt{d}$, and $a = \lambda + dp(1-p)$. Then $F = dp(1-p)bb^\mathsf{T}$, $\mathsf{A}b = ab$, $\mathsf{A}^{-1}b = b/a$. Orthogonal decomposition gives $\|\widehat{v} - b/a\|_\mathsf{A}^2 \ge a|b^\mathsf{T}\widehat{v} - 1/a|^2$, whereas $\|b/a\|_\mathsf{A}^2 = 1/a$. The assumed relative error bound therefore implies $|ab^\mathsf{T}\widehat{v} - 1| \le \varepsilon$, and $db^\mathsf{T}\widehat{v}$ estimates $d/a = G_{\lambda/d}(p)$ to relative error ε. Apply Proposition E.2 at λ/d, which proves the lower bound in Proposition 3.6. Each vector observation contains one Bernoulli draw. The dimension factor comes from the allowed norm $\sqrt{d}$ and is already present in this rank-one construction. □

Bias and perturbation. The sampling guarantee above controls error relative to the damped natural gradient. Comparing with the undamped natural gradient also requires accounting for the bias introduced by damping. For positive definite F and $\lambda > 0$, choose an orthonormal eigenbasis and write $Fu_j = \mu_j^{\mathrm{eig}} u_j$, $b_j = u_j^{\mathsf{T}} b$, $v = F^{-1}b$ and $v_\lambda = (F + \lambda I)^{-1} b$, with $b \neq 0$. Here $\|z\|_F^2 = z^{\mathsf{T}} F z$. Eigenbasis expansion gives the damping bias

$$\frac{\|v_\lambda - v\|_F^2}{\|v\|_F^2} = \frac{\sum_j (b_j^2/\mu_j^{\mathrm{eig}})[\lambda/(\mu_j^{\mathrm{eig}} + \lambda)]^2}{\sum_j b_j^2/\mu_j^{\mathrm{eig}}}. \tag{A68}$$

The bias depends on the gradient components in the eigenbasis and on the eigenvalues of F. For $\mathsf{A} = F + \lambda I$, let $\widehat{v}_\lambda = (\mathsf{A} + \Delta F)^{-1} b$ be the direction computed from the perturbed Fisher matrix. If $r_{\mathrm{pert}} = \|\mathsf{A}^{-1}\|_2 \|\Delta F\|_2 < 1$, the identity $\widehat{v}_\lambda - v_\lambda = -(I + \mathsf{A}^{-1}\Delta F)^{-1} \mathsf{A}^{-1} \Delta F v_\lambda$ and the Neumann series yield

$$\frac{\|\widehat{v}_\lambda - v_\lambda\|_2}{\|v_\lambda\|_2} \leq \frac{\kappa_2(\mathsf{A})}{1 - r_{\mathrm{pert}}} \frac{\|\Delta F\|_2}{\|\mathsf{A}\|_2}. \tag{A69}$$

Damping reduces sensitivity to matrix error (Martens, 2020), although attaining the required matrix accuracy still depends on sampling variance and spectral scale. These error bounds for estimated directions do not control finite-step KL.

E.2 Resolution under fixed damping

When damping dominates the scalar Fisher information of an event indicator, the reciprocal of the damped information is close to the known value $1/\lambda$. The next result identifies when this approximation attains the required relative accuracy.

Proposition E.4 (Resolution under fixed damping)**.** *Let P be a probability law and A a known measurable event with unknown mass $p = P(A) \in (0, 1/2]$. Set $I = p(1-p)$, fix a known $\lambda > 0$, and let $0 < \rho \leq 1$ and $0 < \delta < 1$. Define $\mathcal{G}_{A,\lambda} := [I + \lambda]^{-1}$. If $I \leq \rho\lambda$, then $1/[(1+\rho)\lambda] \leq \mathcal{G}_{A,\lambda} \leq 1/\lambda$, and $1/\lambda$ has relative error at most ρ. Also, $p \leq \rho\lambda$ implies $I \leq \rho\lambda$, and $I \leq \rho\lambda$ implies $p \leq 2\rho\lambda$. Given $m \geq 1$ independent draws $S_1, \ldots, S_m \sim P$, set $\widehat{p} = m^{-1} \sum_i \mathbf{1}_A(S_i)$ and $\widehat{\mathcal{G}}_{A,\lambda} = [\widehat{p}(1-\widehat{p}) + \lambda]^{-1}$. This estimate has relative error at most ρ with probability at least $1 - \delta$ whenever $m \geq 10 \log(2/\delta)/(\rho^2\lambda)$.*

Proof of Proposition E.4. First check the deterministic approximation. For the sample estimate, choose an error tolerance for the estimated event probability that guarantees the required relative error after inversion, then apply Bernstein's inequality. If $I \leq \rho\lambda$, then $\lambda \leq I + \lambda \leq (1+\rho)\lambda$, and the relative error of $1/\lambda$ is exactly I/λ. The implications for event mass follow from $p/2 \leq I \leq p$. Set $\tau_{\mathrm{dev}} = \rho(I + \lambda)/(1 + \rho)$. Since $p(1-p)$ is 1-Lipschitz, on $|\widehat{p} - p| \leq \tau_{\mathrm{dev}}$ the reciprocal estimate has relative error at most $\tau_{\mathrm{dev}}/(I + \lambda - \tau_{\mathrm{dev}}) = \rho$. Bernstein for bounded centered indicators of variance I (Tropp, 2012) gives

$$\mathbb{P}\{|\widehat{p} - p| > \tau_{\mathrm{dev}}\} \leq 2e^{-m\tau_{\mathrm{dev}}{}^2/(2I + 2\tau_{\mathrm{dev}}/3)} \leq 2e^{-3m\rho^2\lambda/28}. \tag{A70}$$

The last bound uses $\rho(I + \lambda)/2 \leq \tau_{\mathrm{dev}} \leq (I + \lambda)/2$ and $2I + 2\tau_{\mathrm{dev}}/3 \leq 7(I + \lambda)/3$. Since $10 > 28/3$, the stated sample count makes the probability at most δ. □

F Population-KL certification for finite candidate sets

Section 3.4 summarizes the following guarantee for affine classifiers. We prove it below and then extend it to gain comparisons and to sequential validation of proposed updates.

Proposition F.1 (Population-KL certification)**.** *Fix a softmax classifier with affine logits $w_a^{\mathsf{T}} z(x)$ on a finite class set and an update direction with rows u_a. Include any bias coordinate in $z(x)$, and let $\pi_t(\cdot \mid x)$ have logits $(w_a + t u_a)^{\mathsf{T}} z(x)$. Set the per-input cost $k_t(x) = \mathrm{KL}(\pi_0(\cdot \mid x) \| \pi_t(\cdot \mid x))$. For integers $N_{\mathrm{cand}} \geq 1$ and $m \geq 2$, fix candidate steps $t_0 = 0, t_1, \ldots, t_{N_{\mathrm{cand}}} \geq 0$. Let $X_1, \ldots, X_m$ be draws from a finite reference law Q, independent of the classifier, direction and candidates. Let $\eta > 0$, $0 < \delta < 1$, and suppose known constants $\mathcal{M}_j \geq 0$ satisfy $0 \leq k_{t_j}(x) \leq \mathcal{M}_j$ on the support of Q. If $\|z(x)\|_2 \leq R_{\mathrm{feat}}$ on this support for known $R_{\mathrm{feat}} \geq 0$, one valid choice is $\mathcal{M}_j = \min\{R_j, R_j^2/8\}$,*

where $R_j = t_j R_{\mathrm{feat}} \max_{a,b} \|u_a - u_b\|_2$. *For* $1 \le j \le N_{\mathrm{cand}}$, *let* $\widehat{\mathcal{K}}_j$ *and* $\widehat{V}_j$ *be the sample mean and unbiased sample variance of* $k_{t_j}(X_i)$. *Set*

$$\mathcal{U}_j = \min\left\{\mathcal{M}_j,\ \widehat{\mathcal{K}}_j + \sqrt{\frac{2\widehat{V}_j \log(2N_{\mathrm{cand}}/\delta)}{m}} + \frac{7\mathcal{M}_j \log(2N_{\mathrm{cand}}/\delta)}{3(m-1)}\right\}, \qquad \mathcal{U}_0 = 0. \tag{A71}$$

Any measurable choice of $\widehat{j}$ *among candidates with* $\mathcal{U}_{\widehat{j}} \le \eta$, *including one that uses the calibration data, satisfies* $\mathbb{E}_Q k_{t_{\widehat{j}}} \le \eta$ *with probability at least* $1 - \delta$. *The zero step is always accepted.*

To certify updates from sampled KL values, we bound the cost of each candidate separately and require all bounds to hold together. A bounded per-input KL cost permits empirical Bernstein certification of a finite set of candidate updates (Maurer & Pontil, 2009). Conditioning on an independent estimation batch fixes the direction and candidate set before calibration.

Consider a softmax model with affine logits and a reference law Q. Let u_a be the direction's row for class a, including any fixed zero reference row. Suppose the input representation satisfies $\|z(x)\|_2 \le R_{\mathrm{feat}}$ on the reference support, with known $R_{\mathrm{feat}} \ge 0$. At step $t \ge 0$, write $d_a = t u_a^{\mathsf{T}} z(x)$, $\pi_a = \pi_\theta(a \mid x)$, and

$$R_{\mathrm{logit}}(t) = t R_{\mathrm{feat}} \max_{a,b} \|u_a - u_b\|_2, \qquad \mathcal{M}(t) = \min\{R_{\mathrm{logit}}(t), R_{\mathrm{logit}}(t)^2/8\}. \tag{A72}$$

The categorical KL from the current model to the updated model obeys

$$k_t(x) = \log \sum_a \pi_a e^{d_a} - \sum_a \pi_a d_a, \qquad 0 \le k_t(x) \le \mathcal{M}(t). \tag{A73}$$

We call $\mathcal{M}(t)$ the per-input KL envelope. This bound on the KL cost at every input is known before sampling. To verify this bound, note that the logit increments have range at most $R_{\mathrm{logit}}(t)$. Jensen's inequality proves nonnegativity, while bounding the log-sum by the largest increment gives $k_t \le R_{\mathrm{logit}}(t)$. For the quadratic bound, the second derivative of $s \mapsto \log \sum_a \pi_a e^{s d_a}$ is the tilted variance of d_a. This variance is at most $R_{\mathrm{logit}}(t)^2/4$, since each increment's squared distance from the midpoint of the range is at most that value. Integrating the second derivative against $1 - s$ on $[0, 1]$ gives $k_t \le R_{\mathrm{logit}}(t)^2/8$.

Proposition F.2 (Finite-candidate certification). *Fix* $0 < \delta < 1$, $\eta > 0$ *and an integer* $N_{\mathrm{cand}} \ge 1$. *Let* $\mathcal{G}$ *be a* σ*-algebra with respect to which a current conditional model* π_θ, *a direction* u, *nonnegative steps* $t_0 = 0, t_1, \ldots, t_{N_{\mathrm{cand}}}$ *and constants* $\mathcal{M}_j \ge 0$ *are measurable. Let* $m \ge 2$ *independent draws from a finite reference law* Q *be independent of* $\mathcal{G}$, *and suppose that*

$$0 \le k_{t_j}(x) := \mathrm{KL}(\pi_\theta(\cdot \mid x) \| \pi_{\theta + t_j u}(\cdot \mid x)) \le \mathcal{M}_j \tag{A74}$$

on the support of Q. *For each candidate with a positive bound on its per-input KL cost, let* $\widehat{\mathcal{K}}_j$ *and* $\widehat{V}_j$ *be the sample mean and unbiased sample variance of its measurable per-input KL costs, and set*

$$\begin{aligned} r_j &= \sqrt{\frac{2\widehat{V}_j \log(2N_{\mathrm{cand}}/\delta)}{m}} + \frac{7\mathcal{M}_j \log(2N_{\mathrm{cand}}/\delta)}{3(m-1)}, \\ \mathcal{U}_j &= \min\{\mathcal{M}_j, \widehat{\mathcal{K}}_j + r_j\}. \end{aligned} \tag{A75}$$

Set $r_0 = \mathcal{U}_0 = 0$, *and* $r_j = \mathcal{U}_j = 0$ *when* $\mathcal{M}_j = 0$. *Choose measurably from the candidates satisfying* $\mathcal{U}_j \le \eta$, *using any or all calibration observations. The selected candidate has* $\mathbb{E}_Q k_{t_j} \le \eta$ *with conditional probability at least* $1 - \delta$ *given* $\mathcal{G}$, *almost surely, and hence with probability at least* $1 - \delta$. *The zero candidate makes this set nonempty.*

Proof of Proposition F.2. A simultaneous upper bound permits selection after observing the calibration data, provided the candidates are fixed before sampling. The per-input cost $k_t(x)$ is continuous in the model parameters, the direction and t, so every $k_{t_j}(X_i)$, $\mathcal{U}_j$ and $\mathbb{E}_Q k_{t_j}$ is a random variable. Because the draws are independent of $\mathcal{G}$, their conditional law given $\mathcal{G}$ is that of m independent draws from Q, while the candidates and constants are fixed. We prove the bound under this conditional law. The failure-probability bound is uniform in the conditioning, so taking expectations gives the unconditional statement.

For $\mathcal{M}_j > 0$, apply the empirical Bernstein bound of Maurer & Pontil (2009) to $k_{t_j}/\mathcal{M}_j$ with failure probability δ/N_{cand}. Rescaling its mean and unbiased variance gives $\mathbb{E}_Q k_{t_j} \le \widehat{\mathcal{K}}_j + r_j$. The deterministic inequality $\mathbb{E}_Q k_{t_j} \le \mathcal{M}_j$ justifies the minimum in $\mathcal{U}_j$. A union bound gives these inequalities for every candidate on one event of probability at least $1-\delta$. Candidates whose bounds are zero have zero KL. On the simultaneous event, every accepted candidate is feasible, regardless of how it is selected. □

Proof of Proposition F.1. The draws are independent of the classifier, direction and candidates, so let $\mathcal{G}$ be the σ-algebra these generate. Since $R_j = R_{\mathrm{logit}}(t_j)$, Equation (A73) shows that $\mathcal{M}_j = \min\{R_j, R_j^2/8\}$ is a valid bound. With these inputs, Equation (A75) coincides with Equation (A71), and Proposition F.2 gives the claim. □

Under the assumptions of Proposition F.2, suppose the query gain G_j is known exactly. The rule may maximize it over accepted candidates. A comparison with the finite-grid optimum follows by replacing $\log(2N_{\mathrm{cand}}/\delta)$ by $\log(4N_{\mathrm{cand}}/\delta)$ in r_j. Applying the same bound to k_{t_j} and $\mathcal{M}_j - k_{t_j}$ gives $|\widehat{\mathcal{K}}_j - \mathcal{K}_j| \le r_j$ simultaneously with probability at least $1-\delta$, where $\mathcal{K}_j = \mathbb{E}_Q k_{t_j}$. Thus every candidate with $\mathcal{K}_j + 2r_j \le \eta$ is accepted, and the selected gain is at least the best gain among those candidates.

Independent validation of proposed updates. Fix an overall failure probability $0 < \delta < 1$. The same bound can validate a proposal chosen by an empirical margin rule. At validation attempt $r \ge 1$, construct the direction and one proposed step from the training history and any additional data reserved for constructing the proposal. Before validation, fix a known upper bound on the proposal's per-input KL cost and a sample size of at least two.

Then draw a fresh independent batch from the finite reference law. Apply Equation (A75) with $N_{\mathrm{cand}} = 1$ and $\delta_r = \delta/[r(r+1)]$. Accept the proposal if its upper bound is at most the current KL budget, and otherwise take the zero step. Every new or revised proposal uses the next index r and a fresh independent batch, including after a rejection.

Let $\mathcal{G}_r$ be the σ-algebra generated by everything observed before batch r, including whether attempt r is made. The proposal, its KL bound, the sample size and the decision to attempt are $\mathcal{G}_r$-measurable. Batch r, conditional on $\mathcal{G}_r$, consists of independent draws from the reference law. Let B_r be the event that attempt r is made and accepts an update that violates its population-KL budget. Proposition F.2 with $\mathcal{G} = \mathcal{G}_r$ gives $\mathbb{P}(B_r \mid \mathcal{G}_r) \le \delta_r$ almost surely, so $\mathbb{P}(B_r) \le \delta_r$ by the tower property. A union bound over r gives

$$\mathbb{P}\{\text{some accepted update violates its population-KL budget}\} \le \sum_{r=1}^{\infty} \frac{\delta}{r(r+1)} = \delta. \tag{A76}$$

Because the decision to attempt is $\mathcal{G}_r$-measurable, this also permits stopping after any completed attempt. Each validation batch counts toward the total sampling budget. Validating one proposal at a time removes the multiple-candidate correction, but it keeps the range term and uses a smaller failure probability at later updates. The sequential guarantee controls KL feasibility. Objective gain is a separate criterion.

G Reproducibility Details

Reporting conventions. Pointwise percentile bootstrap intervals (Davison & Hinkley, 1997) resample queries for learned-head studies, preserving paired method comparisons. In the logistic direction and scalar damping studies, intervals for means use replicate standard errors with a normal approximation. Intervals for success rates use Wilson's method (Wilson, 1927). The grids, endpoints, seeds and confidence rules in Appendices H.2 and I.2 were fixed before evaluation.

Computational environments. Learned-head experiments used Python 3.12 with PyTorch 2.8 and CUDA 12.6, torchvision 0.23, Transformers 4.57, NumPy 2.3 and SciPy 1.16 on NVIDIA RTX 3090 and RTX A6000 GPUs. The logistic, damping and count studies used Python 3.14 with NumPy 2.4 and SciPy 1.18 on an Apple M2 CPU. The text head uses the checkpoint `distilbert/distilbert-base-uncased-finetuned-sst-2-english` at revision `714eb0fa89d2`

`f80546fda750413ed43d93601a13` and the `sst2` split of `nyu-mll/glue` at revision `bcdcba79d07bc864c1c254ccfcedcce55bcc9a8c`.

Randomization. Every study uses fixed seeds. Each learned-head draw comes from a NumPy seed sequence keyed by the study seed, model, query, repetition and sampling stream, and samples at smaller budgets are prefixes of those at larger budgets.

Numerical solvers. Learned-head Fisher matrix and KL computations use float64, and conjugate gradients must reach relative residual at most 10^{-8} within 2,000 iterations. Scalar searches include the zero step, expand a bracket and bisect the KL boundary, so the query gain may peak before the boundary. Feasibility uses tolerance $10^{-7}\eta + 10^{-12}$, and margin-experiment gains use 10^{-7} relative plus 10^{-10} absolute tolerance.

H Count-based calibration experiments

We vary rarity and gain requirements in the four-state family and the event-tilt construction to examine the sample scales in the calibration bounds. Simulated counts represent independent observations. The methods receive the counts and the information specified below. Implementation and reporting conventions appear in Appendix G.

H.1 Calibration in the four-state family

Table A1: Joint success over 2,048 paired count simulations per setting. (a) Four-state law, $\omega = 1.25$, $p_{\mathrm{M}} = 2^{-30}$, $\varepsilon = 1/8$. (b) Event-tilt confidence rule, 95% event-mass confidence set, $\varepsilon = 1/32$, $mp\varepsilon^2 = 16$.

(a) Four-state calibration			**(b) Binary calibration**	
$mp_{\mathrm{M}}\varepsilon^2$	Plug-in (%)	Conservative (%)	p	Confidence (%)
2^4	48.97	52.78	10^{-6}	95.80
2^{12}	48.05	93.90	10^{-5}	96.39
2^{20}	50.29	100.00	10^{-4}	96.29
			2^{-11}	97.07

We compare conservative and plug-in calibration on the four-state law in Equation (A2). Supplying p_{M}, the initial gradient, scalar Fisher information and natural gradient isolates the remaining calibration problem. Write T_ω for the unique boundary with $\mathcal{K}_\omega(T_\omega) = 1/10$, and denote this KL budget by η. The parameter grid is

$$p_{\mathrm{M}} \in \{2^{-30}, 2^{-32}\}, \qquad \omega \in \{1, 1.25, 1.5\}, \qquad \varepsilon \in \{1/8, 1/16\}. \tag{A77}$$

We use $\delta = 0.05$, KL budget $1/10$, and $mp_{\mathrm{M}}\varepsilon^2 = 2^k$ for $k = -2, 0, 2, \ldots, 20$. Each setting has 2,048 repetitions. We generate $n_3 \sim \mathrm{Binomial}(m, \omega p_{\mathrm{M}})$ directly. This yields the count of the state $h = 3$ without generating the m base draws. The rules use only this count, although it is not a sufficient statistic for the full family, and each trial is charged m observations, up to $m = 2^{60}$. Both sampled rules use the same count in each repetition. The conservative rule from the theorem uses

$$\omega_U = \min\{3/2, \max\{1, n_3/(mp_{\mathrm{M}}) + \varepsilon/100\}\} \tag{A78}$$

and returns $\lfloor N_{\mathrm{round}} T_{\omega_U} \rfloor / N_{\mathrm{round}}$, where $N_{\mathrm{round}} = \lceil 100/\varepsilon \rceil$. The plug-in rule returns $T_{\widehat{\omega}}$, where $\widehat{\omega}$ truncates $n_3/(mp_{\mathrm{M}})$ to $[1, 3/2]$, and the oracle returns T_ω. True finite KL and gain are evaluated under the actual law. The plotted joint-success indicator allows 10^{-12} absolute tolerance for KL feasibility and for the gain ratio. Pointwise 95% Wilson intervals quantify Monte Carlo uncertainty.

Figure A2 shows the interior setting $\omega = 1.25$, $\varepsilon = 1/8$, fixed in advance. Table A2 summarizes every grid budget. At $mp_{\mathrm{M}}\varepsilon^2 = 2^{20}$, the conservative rule succeeds in $2048/2048$ trials at every family setting (pointwise Wilson lower endpoint 99.81%). The plug-in rule ranges from 49.32% to 51.71% for $\omega \in \{1.25, 1.5\}$ and attains 100% at $\omega = 1$. The oracle succeeds throughout. At smaller budgets, the conservative rule can fail in a substantial fraction of trials. Its uniform 95% theorem

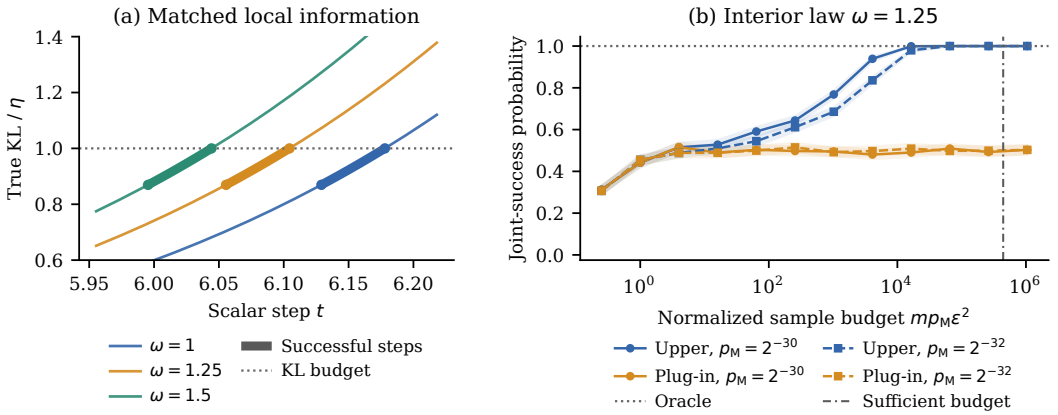

Figure A2: Calibration in the four-state family, whose laws share their initial gradients and scalar Fisher information. (a) True finite KL at $p_{\mathrm{M}} = 2^{-30}$. Thick segments are feasible and attain at least $7/8$ of optimal gain. The $\omega = 1$ and $\omega = 1.5$ segments are disjoint. (b) Paired rules based on counts of $h = 3$ at $\omega = 1.25$, $\varepsilon = 1/8$, with 2,048 binomial simulations per point and pointwise 95% Wilson intervals (Wilson, 1927). The conservative rule, labeled Upper, rounds steps downward. The dash-dot line marks the sample budget sufficient for its theorem guarantee, which does not apply at smaller budgets. Appendix H.1 gives the protocol and ranges over the full grid.

Table A2: All normalized budgets in the four-state study. Each success range spans the 12 settings of $(p_{\mathrm{M}}, \omega, \varepsilon)$: two scale values, three law coefficients and two gain targets. Upper denotes the conservative rule. The last two columns give the largest rates of true KL exceeding 1.1η across these settings. Entries are observed percentages over 2,048 repetitions per setting.

$mp_{\mathrm{M}}\varepsilon^2$	Upper success	Plug-in success	Upper max. rate	Plug-in max. rate
2^{-2}	24.02–73.19	24.02–73.19	40.09	40.09
2^{0}	40.87–89.31	40.87–89.50	19.97	19.97
2^{2}	47.61–99.17	46.78–99.32	3.81	4.25
2^{4}	50.78–100.00	48.93–100.00	0.00	0.05
2^{6}	53.12–100.00	49.95–100.00	0.00	0.00
2^{8}	55.71–100.00	48.10–100.00	0.00	0.00
2^{10}	60.74–100.00	47.80–100.00	0.00	0.00
2^{12}	72.80–100.00	48.05–100.00	0.00	0.00
2^{14}	85.79–100.00	48.78–100.00	0.00	0.00
2^{16}	99.12–100.00	48.29–100.00	0.00	0.00
2^{18}	100.00–100.00	49.02–100.00	0.00	0.00
2^{20}	100.00–100.00	49.32–100.00	0.00	0.00

Table A3: Paired offset/rounding ablation at $\omega = 1.25$, $\varepsilon = 1/8$, reusing the 2,048 counts per row. P is plug-in, O adds the upward offset, and R rounds the step downward. The four central columns give joint-success percentages. The last two give the mean of $\max\{\mathcal{K}/\eta - 1, 0\}$ over all trials, as a percentage of the KL budget.

p_{M}	$mp_{\mathrm{M}}\varepsilon^2$	P	P+R	P+O	P+O+R	P excess	P+O+R excess
2^{-30}	2^{0}	44.14	44.14	44.14	44.14	4.603	4.457
2^{-30}	2^{4}	48.97	51.56	50.39	52.78	1.071	0.935
2^{-30}	2^{8}	49.76	59.47	55.37	64.36	0.263	0.152
2^{-30}	2^{12}	48.05	83.01	70.61	93.90	0.068	0.010
2^{-30}	2^{16}	50.93	100.00	99.02	100.00	0.016	0.000
2^{-30}	2^{20}	50.29	100.00	100.00	100.00	0.004	0.000
2^{-32}	2^{0}	45.51	45.51	45.51	45.51	4.511	4.366
2^{-32}	2^{4}	48.93	48.93	50.00	50.98	1.097	0.962
2^{-32}	2^{8}	51.46	55.47	56.54	61.13	0.260	0.153
2^{-32}	2^{12}	49.71	65.82	70.56	83.59	0.068	0.012
2^{-32}	2^{16}	49.80	95.12	98.58	100.00	0.017	0.000
2^{-32}	2^{20}	50.29	100.00	100.00	100.00	0.004	0.000

guarantee applies only when $mp_{\mathrm{M}}\varepsilon^2 \geq 120000 \log 40 \approx 442666$, which only the last grid budget exceeds.

Separating the offset from rounding. A post hoc ablation reuses the same counts in a two-by-two comparison. One factor adds or omits the positive offset $\varepsilon/100$ to the estimate of ω. The other includes or omits downward rounding of the step with $N_{\mathrm{round}} = \lceil 100/\varepsilon \rceil$. Clipping, boundary search and evaluation remain fixed. Table A3 reports this interior setting. At normalized budget 2^{12}, every rule meets the gain target in every trial for both masses, so the success differences measure feasibility. Both modifications contribute, and the different rounded rates show sensitivity to the location of the true boundary within the rounding grid. At smaller budgets, rounding can instead reduce joint success by moving a step below the gain threshold. The largest observed decrease is 1.86 percentage points across the full grid. At the largest budget, the offset alone and the full rule succeed in every trial across all 12 law/accuracy settings, whereas rounding alone ranges from 77.44% to 100%. For the two interior settings in the table, the unmodified plug-in succeeds in 50.29% of trials, while its mean positive KL excess is only about 0.004% of the budget. Every method meets the gain target there, and none exceeds 1.1η. Strict boundary crossings can thus persist as their size shrinks.

H.2 Event tilts: rarity and required gain

We vary event rarity and the allowed fractional gain loss ε to separate their contributions to sample cost. The grid uses

$$p \in \{10^{-6}, 10^{-5}, 10^{-4}, 2^{-11}\}, \qquad \varepsilon \in \{1/4, 1/8, 1/16, 1/32\}. \tag{A79}$$

Normalized sample budgets $mp\varepsilon^2$ double from $1/4$ to 64, with 2,048 independent simulations of the cell counts per setting. The evaluator uses p to set m, while the rules receive only counts and m. At KL budget 0.1, we compare empirical 90%-budget calibration, four-cell confidence calibration and the direct event-count rule. Both confidence rules use 95% coverage. In the four-cell control, the event and its complement each split into two equal cells, and the confidence set in (A82) for the four cell probabilities yields an interval for the combined probability of the two event cells. Each confidence rule maximizes KL over its entire event-probability interval, including any interior maximizer. All rules restrict the tilt parameter to at most 64 and use relative feasibility tolerance 10^{-7}.

At $p = 10^{-5}$ and $mp\varepsilon^2 = 16$, the event-count confidence rule achieves 96.4–99.1% joint success over $\varepsilon \in \{1/4, 1/8, 1/16, 1/32\}$ (Figure A3), with expected event counts from 256 to 16,384. A smaller gain tolerance therefore requires more expected rare events at the same normalized sample budget.

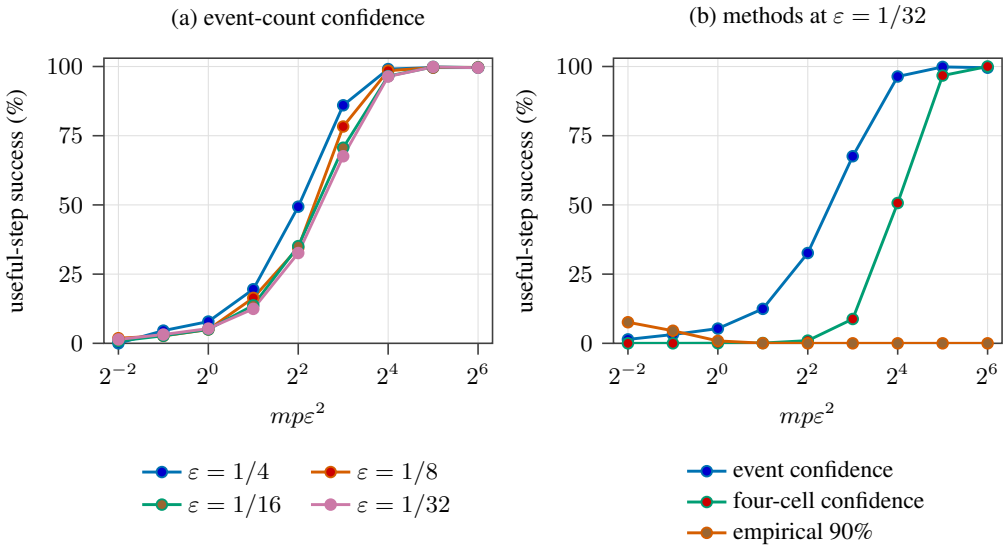

Figure A3: Dependence on the required improvement at $p = 10^{-5}$, $\eta = 0.1$. The two confidence-based rules use 95% confidence. Each point summarizes 2,048 simulated count vectors. Usefulness requires feasibility and a fraction $1 - \varepsilon$ of the reference optimum. At a fixed horizontal coordinate, the number of observations varies with ε. Empirical calibration at 90% of the budget can limit gain at stringent targets.

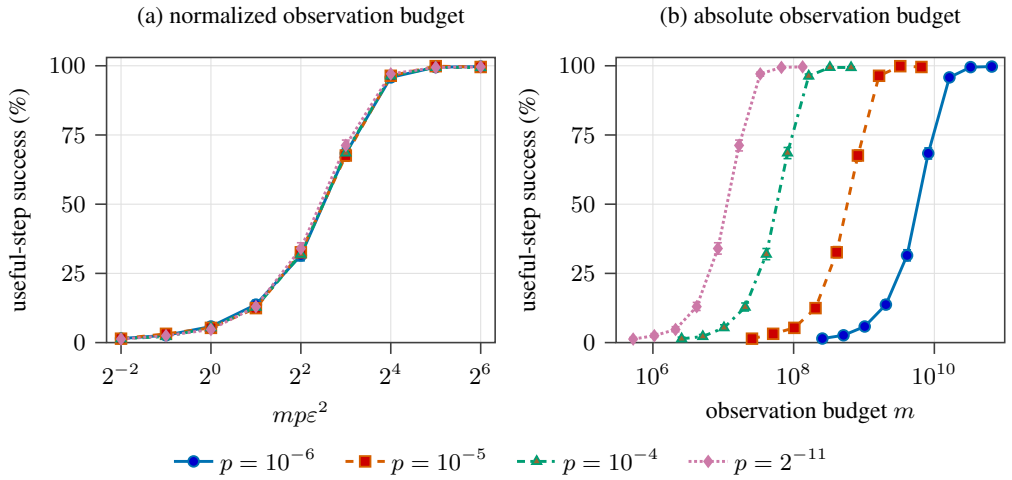

Figure A4: **Calibration cost across event rarities.** The event-count rule uses 95% confidence for event mass, $\varepsilon = 1/32$ and KL budget $\eta = 0.1$. Success requires KL feasibility and at least $31/32$ of optimal event gain. Both panels show the same 2,048-repetition estimates with pointwise 95% Wilson intervals (Wilson, 1927). Multinomial counts simulate m independent observations.

Figure A4 compares the four event masses on normalized and absolute sample scales. At $mp\varepsilon^2 = 16$ and $\varepsilon = 1/32$, the event-count rule succeeds in 95.8% to 97.1% of trials across rarities, at sample sizes from 3.36×10^7 to 1.64×10^{10}. High success rates already appear below the explicit sufficient threshold in Equation (A40).

I LOGISTIC CONTROLS AND DAMPED ESTIMATION

We first compare direction accuracy with finite gain under exact KL calibration. We then measure the effect of estimating KL from samples and how damping changes the required sample count. Finally, we extend the direction-recovery construction to higher dimensions.

I.1 FINITE-STEP IMPROVEMENT WITH A SAMPLED FISHER MATRIX

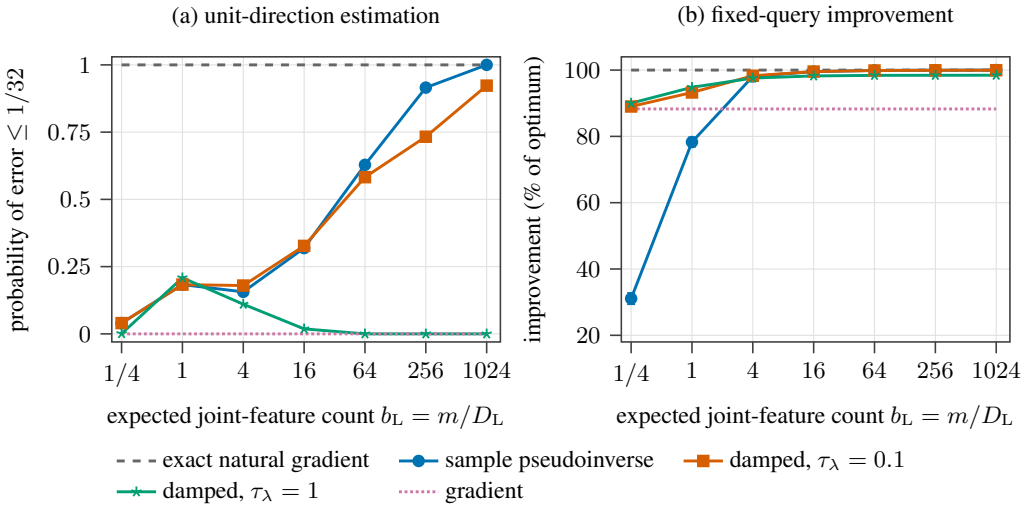

Figure A5: Logistic query updates with $K = 1$, $D_L = 65{,}536$, and conditional KL budget $0.05/D_L$. Left: fraction of directions with Euclidean error at most $1/32$. Right: mean log-likelihood improvement as a percentage of the global constrained optimum, with 95% intervals over 2,048 independent repetitions. Every nonzero direction uses the same exact reference-KL line search. Damping is $\lambda = \tau_\lambda/(4D_L)$.

For the logistic law in Equation (4), we use $K \in \{1, 2\}$, $D_L \in \{256, 4096, 65536\}$, and $b_L = m/D_L \in \{1/4, 1, 4, 16, 64, 256, 1024\}$. The 2,048 independent multinomial count vectors per setting are shared across methods and KL budgets. Each vector represents m independent draws without generating them individually.

The sampled Fisher information matrix is $\widehat{F} = (4m)^{-1} \sum_i X_i X_i^\mathsf{T}$. At zero this also equals the average outer product of the observed scores, since $(Y - 1/2)^2 = 1/4$. We compare the population natural-gradient direction, $\widehat{F}^\dagger e_1$, $(\widehat{F} + \tau_\lambda I_2/(4D_L))^{-1} e_1$ for $\tau_\lambda = 0.1, 1$, and the gradient e_1. Any positive diagonal preconditioner gives the same ray as the gradient for this query, so the gradient also represents diagonal methods. Nonzero directions are normalized, and a zero pseudoinverse direction gives a zero update. Direction success means Euclidean error at most $1/32$ relative to the normalized population natural-gradient direction.

Query improvement is $J(\theta) = \log 2 - \log(1 + e^{-\theta_1})$. Write $B_K(\theta) = D_L \mathcal{K}_c^L(\theta)$ for the scaled KL in (A54). The scaled budget is $\bar{\eta} = D_L \eta$, where η is the absolute conditional-KL budget. For each ray, we calibrate the step by bisection using the reference KL at $\bar{\eta} \in \{0.005, 0.01, 0.05, 0.5\}$, so the comparison isolates direction quality. Reference calibration receives the true covariate probabilities. The benchmark is $J_K^*(\bar{\eta}) = \max_{B_K(\theta) \le \bar{\eta}} J(\theta)$. To compute this benchmark, note that at fixed

$\theta_1 = t > 0$, strict convexity gives a unique minimizing $\theta^\star_{2,K}(t) \in (-t, 0)$ satisfying

$$K \tanh(\theta^\star_{2,K}(t)/2) + \tanh((t + \theta^\star_{2,K}(t))/2) = 0. \tag{A80}$$

The KL after minimizing over the second coordinate increases strictly to infinity, so a second bisection solves $B_K(t, \theta^\star_{2,K}(t)) = \bar{\eta}$. For $K = 1$, $\theta^\star_{2,1}(t) = -t/2$: the population natural-gradient ray is globally optimal. For $K = 1, D_{\mathrm{L}} = 65536, \bar{\eta} = 0.05, b_{\mathrm{L}} = 1$, the sampled Fisher matrix $\widehat{F}$ is singular in 28.8% of trials. The pseudoinverse, the two damping levels and the gradient achieve mean gain fractions of $78.3, 93.2, 94.8, 88.3\%$, respectively. The corresponding fractions are $90.3, 96.0, 97.4, 94.9\%$ at $K = 2$. The methods have the same ordering at all four budgets. At $K = 1, b_{\mathrm{L}} = 16$, pseudoinverse and lightly damped gains reach 99.5% and 99.6%. Only 31.9% of pseudoinverse directions meet the Euclidean tolerance. Direction success rises to 91.6% at $b_{\mathrm{L}} = 256$. Stronger damping retains its bias, with gains of 98.2% and 98.4% at these two sample budgets. Figure A5 reports uncertainty for these estimates.

Direction alignment in the Fisher metric. We evaluate direction alignment and reference-calibrated gain using the same count vectors. Consider a linear objective with gradient b, Fisher matrix $F \succ 0$ and ray u satisfying $b^\mathsf{T} u > 0$. Under quadratic KL, the optimal gain along this ray divided by the global optimal gain is

$$\frac{b^\mathsf{T} u}{\sqrt{u^\mathsf{T} F u}\sqrt{b^\mathsf{T} F^{-1} b}}. \tag{A81}$$

This is the Fisher cosine between the estimated ray and the population natural gradient. Zero directions contribute zero to both the local and finite-gain means. Table A4 compares Fisher alignment with the Euclidean criterion and actual nonlinear gain at $K = 1$, $D_{\mathrm{L}} = 65536$ and $\bar{\eta} = 0.05$. At $m/D_{\mathrm{L}} = 16$, the mean cosine is 99.46% and mean finite gain is 99.54%, despite the low Euclidean pass rate in Table A4. Across all 840 combinations of method, law, sample budget and KL budget, the largest mean absolute difference between the Fisher cosine and finite-gain fraction is 6.08 percentage points, at $\bar{\eta} = 0.5$.

Table A4: Post hoc Fisher alignment for the pseudoinverse direction at $K = 1$, $D_{\mathrm{L}} = 65536$ and $\bar{\eta} = 0.05$, reusing the 2,048 trials per row. All entries after the sample budget are percentages. Euclidean pass uses the $1/32$ tolerance. Fisher cosine and finite gain are means over trials. S_{90} is the fraction attaining 90% of global optimal finite gain under reference-KL calibration. Zero directions remain in every denominator.

m/D_{L}	Euclidean pass	Fisher cosine	Finite gain	S_{90}
0.25	4.0	30.21	31.08	5.1
1	18.3	77.31	78.30	41.1
4	15.6	97.80	98.10	96.2
16	31.9	99.46	99.54	100.0
64	62.8	99.88	99.89	100.0
256	91.6	99.97	99.97	100.0
1024	100.0	99.99	99.99	100.0

I.2 Sample-based logistic updates

We now estimate the KL cost as well as the direction to measure their combined effect on useful-step selection. We use the fixed query $(e_1, 1)$ and global reference-KL gain benchmark from Appendix I.1. The grid is $K \in \{1, 2\}$, $D_{\mathrm{L}} \in \{4096, 65536\}$, $\bar{\eta} \in \{0.01, 0.05, 0.5\}$, $\eta = \bar{\eta}/D_{\mathrm{L}}$, and $m/D_{\mathrm{L}} \in \{1/4, 1, 4, 16, 64, 256, 1024\}$. Each count setting has 2,048 exact multinomial repetitions shared across methods and budgets. This calibration comparison uses draws independent of those used in the direction-estimation experiment in Appendix I.1. Sample-based methods receive counts, absolute damping and budget values, and the query gradient $e_1/2$. They do not impose the construction's equality between the masses of e_1 and e_2. In this family, the exact direction and encoding size determine the cell masses, but sampled calibration uses only the counts. This isolates the chosen rule's response to KL uncertainty. We compare the population natural-gradient and pseudoinverse directions, damped directions with $\lambda \in \{0.1, 1\}/(4D_{\mathrm{L}})$, and the ordinary gradient. Nonzero directions are

normalized, with zero directions giving zero updates. We apply empirical, reference and confidence calibration to each direction. For counts n_i from m covariate draws, let $\widehat{p}_i = n_i/m$, and let Δ_4 be the simplex on the four covariate cells. Confidence calibration uses

$$\mathcal{P}_m = \{\boldsymbol{w} \in \Delta_4 : m d_{\mathrm{B}}(\widehat{p}_i \| w_i) \le \log(8/\delta) \text{ for all } i\}, \qquad \delta = 0.05. \tag{A82}$$

Two-sided Bernoulli Chernoff bounds give failure probability at most $\delta/4$ per cell (Hoeffding, 1963). A union bound therefore gives $\mathbb{P}(\boldsymbol{p} \in \mathcal{P}_m) \ge 1 - \delta$. Boundary values of d_{B} are defined by limits, so zero and full counts are allowed. The cost $\mathcal{K}_w^{\mathrm{L}}(\theta) = \sum_x w_x \log\cosh(\theta^{\mathsf{T}} x/2)$ is linear in the covariate masses w, so greedy allocation to cells ordered by $|v^{\mathsf{T}} x|$ maximizes it. The cell ordering is unchanged along each ray for $t \ge 0$. For each nonzero direction v, confidence calibration returns the largest $t \in [0, 64]$ satisfying $\max_{w \in \mathcal{P}_m} \mathcal{K}_w^{\mathrm{L}}(tv) \le \eta$. Empirical and reference calibration replace this worst-case KL by the cost under the observed cell frequencies and the true cell masses, respectively. Coverage of the true cell masses guarantees KL feasibility even for directions selected from the same counts, so sample splitting is unnecessary.

All searches use cap $t = 64$ and 50 bisections. Relative feasibility tolerance is 10^{-7}. The primary gain target is 90%, with sensitivity analyses at 75% and 95%.

For confidence calibration in Table A5, joint success rises from 1.1% [0.7,1.6] to 98.9% [98.4,99.3] with population directions as m/D_{L} rises from 16 to 256. Table A6 gives the corresponding pseudoinverse results. For empirical pseudoinverse calibration at ratios $16, 64, 256, 1024$, feasibility is $43.2, 45.8, 47.8, 49.2\%$, while the 95th-percentile KL ratios decrease to $1.411, 1.173, 1.081, 1.039$, respectively. Smaller violations therefore need not imply a high rate of strict feasibility. At $K = 1, \eta = 0.05/D_{\mathrm{L}}$, the gradient ray reaches at most 88.3% of global optimal gain, making the 90% target unattainable on that ray.

I.3 Sample dependence on the damping scale

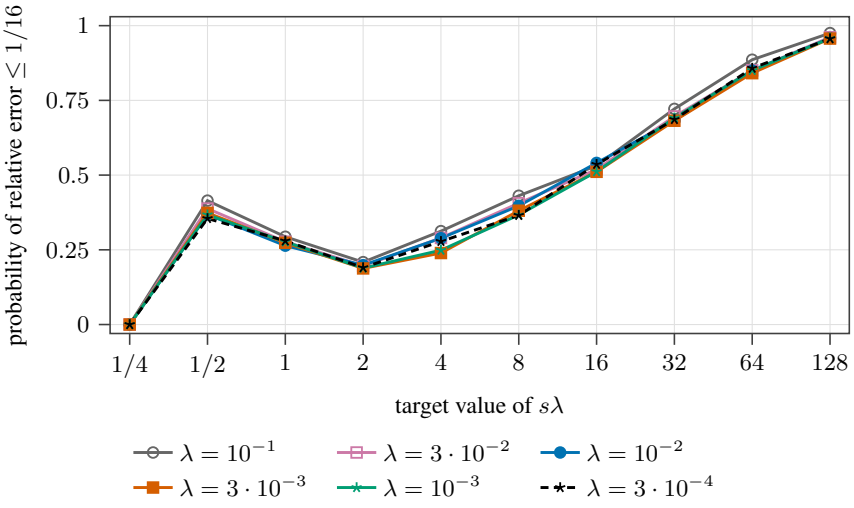

Figure A6: Estimates of the reciprocal of damped scalar Fisher information from paired observations at $p = \lambda$, with 4,096 repetitions per point and $2s$ observations. At small sample sizes, the binomial counts meeting the error threshold change discretely, so success need not increase at every grid point.

We set $p = \lambda$ so that event information and damping have comparable scales, as in the lower-bound construction. For s independent pairs of Bernoulli observations, let N_{disc} count pairs whose outcomes differ. Each pair has discordance probability $2p(1-p)$. Thus

$$N_{\mathrm{disc}} \sim \mathrm{Binomial}(s, 2p(1-p)), \qquad \widehat{F} = N_{\mathrm{disc}}/(2s), \qquad F = p(1-p). \tag{A83}$$

We use $\lambda \in \{0.1, 0.03, 0.01, 0.003, 0.001, 0.0003\}$ and $s\lambda \in \{1/4, 1/2, 1, 2, 4, 8, 16, 32, 64, 128\}$, rounding s to the nearest integer. There are 4,096 exact binomial repetitions per setting, each

Table A5: Direction and calibration controls at $K = 1$, $D_{\mathrm{L}} = 65{,}536$ and $\eta = 0.05/D_{\mathrm{L}}$. The population method uses the exact direction, while the pseudoinverse method estimates it from the counts. Reference calibration uses the true law. Confidence calibration uses the simultaneous confidence set for cell masses from those same counts. All metrics are percentages over 2,048 paired repetitions per row. *Feas.* is true KL feasibility. *Gain* is mean improvement divided by the global constrained query optimum, with infeasible gains set to zero. S_{90} requires both feasibility and at least 90% of that optimum. Brackets are pointwise 95% Wilson intervals for S_{90}.

m/D_{L}	Direction	Calibration	Feas.	Gain	S_{90} [95% CI]
16	Population	Reference	100.0	100.0	100.0 [99.8, 100.0]
16	Population	Confidence	99.9	78.0	1.1 [0.7, 1.6]
16	Pseudoinverse	Reference	100.0	99.5	100.0 [99.8, 100.0]
16	Pseudoinverse	Confidence	99.9	76.9	0.5 [0.3, 1.0]
64	Population	Reference	100.0	100.0	100.0 [99.8, 100.0]
64	Population	Confidence	100.0	88.0	23.6 [21.8, 25.5]
64	Pseudoinverse	Reference	100.0	99.9	100.0 [99.8, 100.0]
64	Pseudoinverse	Confidence	100.0	87.6	20.8 [19.1, 22.7]
256	Population	Reference	100.0	100.0	100.0 [99.8, 100.0]
256	Population	Confidence	100.0	93.6	98.9 [98.4, 99.3]
256	Pseudoinverse	Reference	100.0	100.0	100.0 [99.8, 100.0]
256	Pseudoinverse	Confidence	100.0	93.5	98.7 [98.1, 99.1]

Table A6: Pseudoinverse directions with confidence calibration at $\eta = 0.05/D_{\mathrm{L}}$, over 2,048 repetitions per row. A_{dir} is the percentage with direction error at most $1/32$. Feasibility is evaluated under the true law. Normalized feasible gain and joint success S_{90} use the global query optimum as in Table A5. Brackets are pointwise 95% Wilson intervals for S_{90}.

K	D_{L}	m/D_{L}	A_{dir}	Feas.	Gain	S_{90} [95% CI]
1	4,096	16	33.7	100.0	77.0	0.8 [0.5, 1.3]
1	4,096	64	61.2	100.0	87.5	19.2 [17.6, 21.0]
1	4,096	256	91.8	100.0	93.5	98.5 [97.9, 98.9]
1	65,536	16	33.2	99.9	76.9	0.5 [0.3, 1.0]
1	65,536	64	60.8	100.0	87.6	20.8 [19.1, 22.7]
1	65,536	256	91.6	100.0	93.5	98.7 [98.1, 99.1]
2	4,096	16	39.7	100.0	79.9	1.1 [0.7, 1.7]
2	4,096	64	68.0	100.0	88.9	32.4 [30.4, 34.5]
2	4,096	256	96.2	100.0	94.2	100.0 [99.8, 100.0]
2	65,536	16	39.3	100.0	80.0	1.5 [1.0, 2.1]
2	65,536	64	68.2	100.0	89.0	32.1 [30.1, 34.1]
2	65,536	256	95.1	100.0	94.2	99.9 [99.6, 100.0]

representing $2s$ observations. Success requires relative error in the estimated reciprocal of the damped scalar Fisher information $|(F+\lambda)/(\widehat{F}+\lambda)-1| \le 1/16$. At $s\lambda = 128$, success spans 95.6–97.5% over the six scales, using 2,560–853,334 observations, compared with 84.1–88.6% at 64.

I.4 Direction recovery in higher dimensions

The two-dimensional obstruction persists in higher dimensions, with one active query in each of several blocks. Let $d = 2n_{\mathrm{blk}}$ for an integer $n_{\mathrm{blk}} \ge 1$. A draw selects one block uniformly and places the two-feature covariate there, with zeros elsewhere. All blocks share the same covariate law, and every block has a positive-label query:

$$F_d(0) = \frac{I_{n_{\mathrm{blk}}} \otimes \left(\begin{smallmatrix} K+1 & 1 \\ 1 & K+1 \end{smallmatrix}\right)}{4n_{\mathrm{blk}} D_{\mathrm{L}}}, \qquad \ell_d(\theta) = \frac{1}{n_{\mathrm{blk}}} \sum_j \log \mathrm{sigm}(\theta_1^{(j)}). \tag{A84}$$

The Fisher information matrix has full rank and condition number $1 + 2/K \le 3$. The constraint is ${n_{\mathrm{blk}}}^{-1} \sum_j B_K(\theta^{(j)}) \le \bar{\eta}$, equivalent to conditional KL at most $\bar{\eta}/D_{\mathrm{L}}$. For $\bar{\theta} = {n_{\mathrm{blk}}}^{-1} \sum_j \theta^{(j)}$, convexity of B_K and concavity of $\log \mathrm{sigm}$ give

$$B_K(\bar{\theta}) \le {n_{\mathrm{blk}}}^{-1} \sum_j B_K(\theta^{(j)}), \qquad \log \mathrm{sigm}(\bar{\theta}_1) \ge {n_{\mathrm{blk}}}^{-1} \sum_j \log \mathrm{sigm}(\theta_1^{(j)}). \tag{A85}$$

Repeating the average block therefore preserves feasibility and does not decrease the objective. The global gain optimum is exactly $J_K^*(\bar{\eta})$.

Complexity of the block construction. The shared block law also preserves the direction-recovery obstruction. For the computational construction, let $D_{\mathrm{L}} = 2^{n+5}$. When n_{blk} is a power of two, selecting a block requires $\log_2 n_{\mathrm{blk}}$ additional fair bits, so exact sampling stays polynomial-time. A known projection then recovers the two-dimensional SAT test from any accurate high-dimensional output. Writing $u_K = (K+1, -1)^{\mathsf{T}}/\sqrt{(K+1)^2+1}$, the unit direction is $\nu_d = \mathbf{1}_{n_{\mathrm{blk}}} \otimes u_K/\sqrt{n_{\mathrm{blk}}}$. The unit vector $w = \mathbf{1}_{n_{\mathrm{blk}}} \otimes e_2/\sqrt{n_{\mathrm{blk}}}$ gives $w^{\mathsf{T}}\nu_d = (u_K)_2$, so error $1/32$ preserves the SAT threshold. For rational output, with $S = \sum_j \widehat{\nu}_{2j}$, the test $w^{\mathsf{T}}\widehat{\nu} < -3/8$ is exactly $S < 0,\ 64S^2 > 9n_{\mathrm{blk}}$. The test identifies the unsatisfiable case on every valid output, which gives NP-hardness for fixed or polynomially bounded power-of-two block counts.

For $K = 1, 2$, the $2n_{\mathrm{blk}}$ nonzero coordinate-atom masses each change by $1/(n_{\mathrm{blk}} D_{\mathrm{L}})$ and the zero mass by $2/D_{\mathrm{L}}$. Thus the total variation distance for one draw is $2/D_{\mathrm{L}}$, also with the fair label. Lemma A.1 and the same two-law test give $m \ge D_{\mathrm{L}}/6$.

Sampling experiment. We compare separate estimation in each block with pooling observations under the known common block law. The grid fixes $D_{\mathrm{L}} = 65536, \bar{\eta} = 0.05$, uses $K = 1, 2$, $d \in \{2, 8, 32, 128, 256\}$, and $m/(n_{\mathrm{blk}} D_{\mathrm{L}}) \in \{1/4, 1, 4, 16, 64, 256, 1024\}$, adding fixed-total $m = 16D_{\mathrm{L}}, 256D_{\mathrm{L}}$ comparisons. Each setting has 1,024 paired multinomial repetitions. We compare the population natural-gradient direction, separate and pooled pseudoinverse directions, and the gradient direction. Two further methods use damped directions with $\lambda = \tau_\lambda/(4n_{\mathrm{blk}} D_{\mathrm{L}})$ for $\tau_\lambda = 0.1, 1$. Block magnitudes are retained before global normalization, and zero directions give zero updates. Calibration always uses the reference law.

Table A7 compares increasing dimension at matched per-block sample sizes and at a fixed total budget. At $d = 256, m/(n_{\mathrm{blk}} D_{\mathrm{L}}) = 16$, both values of K have zero direction successes (95% Wilson upper endpoint 0.37%) despite gains of 97.60% and 98.63%. At fixed $m = 16D_{\mathrm{L}}, K = 1$, pooling instead gives 99.55% gain versus 29.59% for separate estimation. The gradient gains remain 88.28% and 94.87% for $K = 1, 2$, independent of dimension.

J Calibration in learned classifier heads

These experiments test whether stopping at a sampled KL boundary is reliable in learned models without constructed rarity. Appendix J.1 fixes the models and queries, Appendix J.2 the directions, calibration rules and outcome measures, Appendices J.3 and J.4 report the full-budget and margin experiments, and Appendix J.5 describes sampling costs and uncertainty.

Table A7: (a) Separate-block pseudoinverse at $m/(n_{\mathrm{blk}}D_{\mathrm{L}}) = 16$. Direction successes are out of 1,024 and gains are percentages of the global optimum with 95% normal half-widths. (b) Gain at fixed total $m = 16D_{\mathrm{L}}, K = 1$. All methods use reference-KL calibration.

(a) Matched expected count per block: $m/(n_{\mathrm{blk}}D_{\mathrm{L}}) = 16$

	$K = 1$		$K = 2$	
d	Successes	Gain (%)	Successes	Gain (%)
2	361	99.53 ± 0.04	410	99.75 ± 0.02
8	0	98.15 ± 0.08	1	98.96 ± 0.04
32	0	97.71 ± 0.05	0	98.68 ± 0.03
128	0	97.62 ± 0.03	0	98.64 ± 0.01
256	0	97.60 ± 0.02	0	98.63 ± 0.01

(b) Fixed total samples: $m = 16D_{\mathrm{L}}$, $K = 1$

d	Separate	$\tau_\lambda = 0.1$	$\tau_\lambda = 1$	Gradient	Pooled
2	99.53	99.54	98.19	88.28	99.53
8	91.18	92.47	95.29	88.28	99.53
32	73.01	50.29	83.26	88.28	99.59
128	43.11	73.64	79.15	88.28	99.54
256	29.59	80.64	82.23	88.28	99.55

J.1 Models and fixed interventions

To study calibration with learned representations, each intervention updates the affine softmax head for one query while keeping the representation fixed. We restore the original head before each query and repetition. Class-stratified round-robin sampling selects 32 initially misclassified queries per model before evaluation. Table A8 lists the reference-set sizes, accuracies on the splits used to select queries, and numbers of free parameters in the heads.

Table A8: Fixed models. Features include an intercept, and the last class's logits are fixed at zero to remove redundant head parameters.

	CIFAR-10 / ResNet-18	SST-2 / DistilBERT
Reference inputs	50,000 training inputs	67,349 training inputs
Query split (accuracy)	Test (95.34%)	Validation (91.055%)
Selected errors	32 of 466	32 of 78
Feature dimension / free head parameters	513 / 4,617	769 / 769

ResNet-18 (He et al., 2016) uses the final checkpoint of a prescribed 200-epoch CIFAR-10 run (Krizhevsky, 2009), without test-based checkpoint selection. Its stem is 3×3, stride one, without max pooling. Training uses SGD (learning rate 0.1, momentum 0.9, weight decay 5×10^{-4}) and a batch size of 128. We use cosine learning-rate decay, random crops with four-pixel padding and horizontal flips. Pixels divided by 255 are normalized by means $(0.4914, 0.4822, 0.4465)$ and standard deviations $(0.2470, 0.2435, 0.2616)$. Features are the 512 final average-pooled values.

DistilBERT (Sanh et al., 2019) uses a fixed GLUE SST-2 checkpoint (Wang et al., 2018; Socher et al., 2013), sequence length 128 and the 768-dimensional representation after the pre-classifier and ReLU. Feature extraction uses evaluation mode without augmentation. The reference law Q is uniform over training inputs. Accuracies on the reference populations are 100% and 98.855%.

J.2 Directions, step selection and outcomes

We compare reference and sampled damped natural-gradient directions with the gradient direction. This isolates how information about the background input law affects the update. All three use

the exact query gradient. For free head parameters θ, put $b = \nabla_\theta \log \pi_\theta(y_* \mid x_*)$ and $s_\theta(y, x) = \nabla_\theta \log \pi_\theta(y \mid x)$.

The model Fisher information matrix and reference direction are

$$F_Q = \mathbb{E}_{x\sim Q} \sum_y \pi_\theta(y \mid x) s_\theta(y, x) s_\theta(y, x)^\mathsf{T}, \tag{A86}$$

$$v_Q = (F_Q + \lambda I)^{-1} b / \|(F_Q + \lambda I)^{-1} b\|_2. \tag{A87}$$

For the estimated direction, we replace the input expectation by a sample average. The gradient direction is $b/\|b\|_2$. Labels of the reference inputs are used only for the prediction diagnostics, not for the direction or KL. These experiments use damped directions without a conditioning restriction. The theoretical construction instead uses an undamped natural gradient and a Fisher matrix with condition number at most 3. For a unit direction v, define

$$G_v(t) = \log \pi_{\theta+tv}(y_* \mid x_*) - \log \pi_\theta(y_* \mid x_*), \tag{A88}$$

$$\mathcal{K}_v(t) = \mathbb{E}_Q \, \mathrm{KL}(\pi_\theta(\cdot \mid x) \| \pi_{\theta+tv}(\cdot \mid x)). \tag{A89}$$

Calibration maximizes the known query gain under sampled or reference KL. Reference directions and calibration require extra population information, whereas empirical calibration uses sampled KL. For each fixed direction, the reference-budget benchmark and success indicator are

$$G_v^*(\eta) = \sup_{t\geq 0:\mathcal{K}_v(t)\leq\eta} G_v(t), \qquad S_\alpha(t) = \mathbf{1}\{\mathcal{K}_v(t) \leq \eta,\ G_v(t) \geq \alpha G_v^*(\eta)\}. \tag{A90}$$

The supremum is finite because the initial query probability is positive. If a ray leaves every reference conditional distribution unchanged, its optimal query gain may be approached only as $t \to \infty$. If at least one reference input has nonconstant class-logit increments, positive baseline softmax probabilities give $\mathcal{K}_v(t) \to \infty$. The feasible interval is then compact and the supremum is attained. We use $\alpha = 0.75$, with 0.90 as a secondary threshold. The benchmark always uses the full budget, including for the 0.9η rule. We also report ratios of population KL to the budget, violations above 1.1η, corrected queries and changes in predictions on the reference population.

Why stopping at an estimated boundary violates the budget about half the time. Strict feasibility is sensitive to selection at an estimated boundary. Fix a ray independently of its calibration data. Suppose $\mathcal{K}_v$ and its empirical estimate $\widehat{\mathcal{K}}_{v,m}$ are continuous and strictly increasing on a fixed neighborhood of a positive root t^* of $\mathcal{K}_v(t) = \eta$. Calibration selects a root $\widehat{t}_m$ of $\widehat{\mathcal{K}}_{v,m}(t) = \eta$, and this root converges in probability to the population boundary. On this neighborhood, the boundary comparison gives

$$\mathcal{K}_v(\widehat{t}_m) > \eta \quad \Longleftrightarrow \quad \widehat{\mathcal{K}}_{v,m}(t^*) < \eta. \tag{A91}$$

If $\sqrt{m}[\widehat{\mathcal{K}}_{v,m}(t^*) - \eta]$ has a centered, nondegenerate normal limit, the strict violation probability tends to $1/2$, even though the violation magnitude vanishes in probability. This concerns exact feasibility before numerical tolerances. Inactive constraints, caps or rays estimated from the calibration data need not satisfy these conditions. The margin rule constrains the empirical finite-step KL, so it guards against sampling error rather than against the error of a quadratic KL approximation.

J.3 Full-budget experiment

We cross three directions and two calibration rules to distinguish direction-estimation effects from calibration effects. The grid is fixed before evaluation. Sample sizes are $m \in \{256, 1024, 4096\}$, with budgets $\eta \in \{0.01, 0.001\}$ and default damping $\lambda = 0.001$. The values $\lambda = 0.0001, 0.01$ are added at $m = 1024, \eta = 0.01$. Fisher matrix estimation and calibration use independent batches sampled with replacement. Corresponding methods use the same sampled inputs, and smaller batches use the first draws of the larger batches. Each condition has 20 repetitions for each of 32 queries. The primary comparison subtracts the raw gain under empirical calibration from that under reference calibration, using the reference direction at $m = 1024, \eta = 0.01, \lambda = 0.001$.

Raw gains include infeasible steps and can exceed the reference-budget optimum. The primary calibration intervals include zero. At the primary setting, empirical calibration along the reference direction produces KL above 1.1η in 27.34% and 37.19% of interventions. The direction comparisons in Table A9 depend on budget and have unequal sampling costs. Joint success is a post hoc measure in this experiment. At the primary setting, every empirical update reaches 90% of its own ray benchmark. Joint success along the reference direction is 47.50% [42.97,51.72] on CIFAR and 45.63% [41.88,49.38] on SST. These failures are KL violations.

Table A9: Paired raw-gain contrasts at $m = 1024$, $\lambda = 0.001$, in 10^{-3} nats, with 95% query-bootstrap intervals. Direction contrasts use reference calibration.

Contrast (budget)	CIFAR-10	SST-2
Reference minus empirical calibration (0.01)	-1.59 [$-5.09, 1.75$]	-0.112 [$-1.214, 0.845$]
Reference direction minus estimated direction (0.01)	2.32 [$-0.36, 5.13$]	0.13 [$-5.69, 4.90$]
Estimated direction minus gradient direction (0.01)	156.97 [71.08,249.32]	1717.06 [1401.41,2021.48]
Reference direction minus estimated direction (0.001)	4.76 [1.10,8.64]	76.58 [52.35,101.32]

J.4 Margin experiment and outcomes

We test whether reserving 10% of the empirical KL budget reduces boundary crossings while retaining useful gain. We fix the grid before evaluation and reuse the checkpoints, queries, reference law, extracted features and reference directions at $\lambda = 0.001$. For each query, we run 20 fresh repetitions at $m \in \{256, 1024, 4096, 16384\}$, using the same sampled inputs to compare calibration rules. The four rules combine empirical or reference KL with the budget η or 0.9η. We apply all four rules at each KL budget. The following lemma bounds the gain lost by reserving budget under the exact law. This gives a benchmark for the empirical margin comparison.

Lemma J.1 (Exact-budget margin for affine-head log-likelihood)**.** *Let $\pi_{\theta+tv}(\cdot \mid x)$ be a softmax classifier whose logits are affine in $t \geq 0$, with fixed parameters θ and unit direction v. Fix a finite reference law Q and a labeled query (x_*, y_*). Define the query log-likelihood gain, population KL cost and ray benchmark by*

$$\begin{aligned} G_v(t) &= \log \pi_{\theta+tv}(y_* \mid x_*) - \log \pi_\theta(y_* \mid x_*), \\ \mathcal{K}_v(t) &= \mathbb{E}_Q \, \mathrm{KL}(\pi_\theta(\cdot \mid X) \| \pi_{\theta+tv}(\cdot \mid X)), \\ G_v^*(\eta) &= \sup_{t \geq 0 \colon \mathcal{K}_v(t) \leq \eta} G_v(t). \end{aligned} \tag{A92}$$

Then $G_v^*(\chi_{\mathrm{bud}}\eta) \geq \chi_{\mathrm{bud}} G_v^*(\eta)$ for every $\eta > 0$ and $0 < \chi_{\mathrm{bud}} < 1$, whether or not either supremum is attained.

Proof of Lemma J.1. The log-partition function is convex, so $\mathcal{K}_v$ is convex and G_v concave, with $\mathcal{K}_v(0) = G_v(0) = 0$.

For every $t \geq 0$ with $\mathcal{K}_v(t) \leq \eta$,

$$\begin{aligned} \mathcal{K}_v(\chi_{\mathrm{bud}} t) &\leq \chi_{\mathrm{bud}} \mathcal{K}_v(t) \leq \chi_{\mathrm{bud}} \eta, \\ G_v(\chi_{\mathrm{bud}} t) &\geq \chi_{\mathrm{bud}} G_v(t). \end{aligned} \tag{A93}$$

Thus $G_v^*(\chi_{\mathrm{bud}}\eta) \geq \chi_{\mathrm{bud}} G_v(t)$ for every such t. Taking the supremum over these candidates proves the claim without assuming that either optimum is attained. □

Both reference rules attain 100% joint success at both gain thresholds, and the smallest reference-margin gain fraction is 96.23%. For empirical calibration at 0.9η, with $\eta = 0.01$ and $m = 16{,}384$, every update meets the 90% gain target. The remaining joint-success failures are KL violations: 3/640 on CIFAR and 40/640 on SST. Of these, zero CIFAR violations and four SST violations exceed 1.1η. At this setting, the mean decline in accuracy on the training reference set stays below 0.15 percentage points under both empirical rules. Compared with full-budget reference calibration, reference-KL margins correct two fewer CIFAR queries and one fewer SST query among 32 targets. Joint success under the empirical margin rule increases at every tested m for both budgets and models.

At $\eta = 0.001$, empirical margin success reaches 98.13% on CIFAR and 91.56% on SST at the largest sample size. The SST interval is [88.13,94.84]. At the 75% gain threshold, all 16 paired differences in joint success between margin and full-budget calibration have positive lower interval endpoints.

Table A10: Results for empirical calibration in Figure 2 at $m = 16{,}384$ and $\eta = 0.01$. Joint success requires KL feasibility and at least 75% of the full-budget ray-optimal gain. Brackets give pointwise 95% query-bootstrap intervals. Margin gain averages each intervention's raw gain as a percentage of the full-budget ray optimum. This mean includes infeasible updates.

	Joint success (%) [95% CI]		
Dataset	Full budget η	Margin 0.9η	Margin gain (%)
CIFAR-10	48.59 $[44.69, 52.50]$	99.53 $[98.91, 100.00]$	99.63
SST-2	50.94 $[47.34, 54.69]$	93.75 $[91.09, 96.25]$	99.79

Coverage is pointwise for each comparison. The 90% target lowers success at some smaller sample sizes.

J.5 Sampling cost and uncertainty

Intervals use query means over 20 repetitions and 5,000 bootstrap resamples of the 32 queries. They quantify variation across the selected queries, conditional on the checkpoints and query selection. An empirical update along the estimated direction uses independent batches of m observations for Fisher matrix estimation and calibration. Gradient and supplied-ray rules use only the calibration batch.